ilmedia

TECHNISCHE UNIVERSITÄT ILMENAU

Knauf, Rainer; Tsuruta, Setsuo; Uehara, Keiichi; Gonzalez, Avelino, J. :

Validation Knowledge Bases and Validation Expert Software Agents : Models of Collective and Individual Human Expertise

Zuerst erschienen als:
Technical Report # 01/04. Tokyo Denki University, School of Information Environment, Chiba New Town, Japan.

Validation Knowledge Bases and Validation Expert Software Agents

Models of Collective and Individual Human Expertise

Technical Report # 01/04

Tokyo Denki University
School of Information Environment
2-1200 Musi-gakuendai, Inzai
Chiba, 270-1382
Japan

Authors:

- **Rainer Knauf**, visiting researcher at TDU
- **Setsuo Tsuruta**, professor of TDU
- **Keiichi Uehara**, researcher of TDU
- **Avelino J. Gonzalez**, temporary visitor at TDU

1. Introduction

Because of the character of their typical application fields, intelligent systems are validated and refined on the basis of human expertise. Experts have different beliefs, experiences, and learning capabilities and are not free of mistakes. Their opinions about the desired system's behavior differ from each other and change over time. Their opinions differ from their previous ones, even in the same context, as a result of misinterpretations, mistakes or new insights. Furthermore, experts are too busy and too expensive to spend that much time for system validation and adjustment. Thus, the experts' workload for system validation is a serious issue.

To make validation results less dependent on the experts' opinions and to decrease the workload of the experts, the importance of storing and using historical validation results / knowledge in a Validation Knowledge Base (***VKB***) was originally proposed in (Tsuruta et al. 2000a) and adopted for a TURING Test (Turing 1950) validation technology in (Knauf et al. 2003).

In the technique described in (Knauf et al. 2002), the result is influenced by the quality of interaction with human experts. Their excessive involvement is both time consuming and costly. In addition, human experts may not always be available or even willing to cooperate, thereby causing delays. In (Tsuruta et al. 2002) this is summarized as *"The bottleneck in acquiring validation knowledge from experts who are busy."*

Since validation is a repeated process and ***VKB***s itself are subject of validation, it might be necessary to urge experts to provide the same knowledge many times. Though intelligent systems must be continually or periodically validated to ensure correctness vis-à-vis the latest findings, it is very unlikely that major changes have to be expected from one validation session to the next for an AI system that runs in long-term practical application.
Therefore, a full-fledged validation effort, including a panel of validation experts, is not constantly required. However, this implies that the knowledge used in validation, namely the set of test cases including their best rated solutions as well as their authors, must persist from one validation exercise to the next. Thus, a way to store, manage, and maintain validation knowledge is required for any practical approach to validation, and this could provide a vehicle for long-term management and improvement of the validation process for intelligent systems.

The validation procedure, as developed so far, covers five steps:

1. test case generation,
2. test case experimentation,
3. evaluation of results,
4. validity assessment, and
5. system refinement.

These steps can be performed iteratively.

Its most expensive part is the test case experimentation, because the test cases have to be solved and rated by both the system under examination and the humans

who perform the examination.[1] This step is especially supported by the ***VKB***. Furthermore, the ***VKB*** is applied to other useful purposes, for example

- to improve the validation methodology itself, e.g. to select experts for the validation panel, and
- to support the identification of an optimal solution among several candidate solutions.

Furthermore, a Validation Expert Software Agent (***VESA***) is developed based on the ***VKB***. A ***VESA*** keeps personal validation knowledge, such as previous validation judgments or the experiences of a human expert. It is an intelligent avatar corresponding to its human origin.
VESAs systematically model human-like validators by keeping personal validation knowledge per corresponding expert. At some point, a ***VESA*** may be able to serve as a substitute for a missing human expert.

The following two sections provide the main ideas of KNAUF's validation framework and TSURUTA's concept of ***VKB*** as developed originally. These sections are followed by the introduction of TSURUTA's and UEHARA's concept of ***VESA*** in the fourth section. After the introduction of all basic concepts, their synergetic usage in an advanced validation framework is described in a fifth section.

Section six is dedicated the prototype experiment that has been performed during Dr. KNAUF's research at Tokyo Denki University. After an introduction of the example domain a formal rule base is derived from the informal domain knowledge.
Based on this knowledge base, all steps of KNAUF's methodology are applied to this example. Of course, the main focus is the prototype Turing test experimentation that includes the new concepts of ***VKB*** and ***VESA***.
In the seventh section, the results of this experiment are derived. In particular, the benefit of ***VKB*** and ***VESA*** is considered in the context of this prototype experiment. Finally, the basic results and insights as well as the derived research focuses are presented in section eight.

2. KNAUF's Validation Framework – An Overview

The validation framework introduced in (Knauf 2000, Knauf et al. 2002) consists of five steps, which can be performed in cycles (see Figure 1):

(1) Test case generation

Here, an appropriate set of test cases **[*TestData, ExpectedOutput*]** is generated. This set meets the competing requirements

- *Coverage* of all possible combinations of inputs which expands the number of test cases to ensure completeness in coverage, and
- *efficiency* which limits the number of test cases to make the process practical.

This step is performed in two sub-steps:

[1] In the process not only the system's solutions, but also the solutions provided by humans are examined. This is performed to estimate the experts' competence for each particular test case.

a) First, a *quasi-exhaustive set of test cases* (***QuEST***) is computed by analyzing the rules and their input/output behavior.
b) Second, the large amount of test cases is limited by utilizing so-called validation criteria. Test cases that don't reach a certain validation necessity degree by considering these criteria will be removed from ***QuEST*** resulting in a *reasonably sized set of test cases* ***ReST***.

A workable compromise between these constraints is central to both the technique developed so far and the improvements reached by introducing the ***VKB***.

(2) Test case experimentation

Intelligent systems emulate human expertise. Therefore, human opinion needs to be considered when evaluating the correctness of a system's response. Through a TURING Test - like validation approach, this step performs a fair evaluation of the correctness and/or dependability of a system's outputs given by imperfect human expertise. It consists of

(1) exercising the set of test data by both the intelligent system and the validating experts and
(2) presenting all results - those provided by the system as well as those provided by the human experts - to the validation panel anonymously.

(3) Evaluation

The third step interprets the results of the experimentation and determines errors attributed to the system and reports it informally.
As a side effect of the previous step, a test case competence assessment of the validators for each particular test case is computed and utilized for a more objective validity statement in the following step.

(4) Validity assessment

In this step, the results of the evaluation are analyzed and conclusions about the system's validity are drawn. Depending on the purpose of the validation statement, the validity is expressed as

- validity degrees associated to test cases,
- validity degrees associated to the system's outputs,
- validity degrees associated to system's rules, and finally
- as a validity degree associated to the entire system.

(5) System refinement

At the first view, the objective of validation is to gain reliable statements on the usefulness and dependability of an intelligent system. In the end, however, we are also interested in developing a more dependable system with a better performance. Therefore, this fifth step, which completes the framework, provides guidance on how to correct or decrease the effects of errors or vulnerabilities detected in the system as a result of the previous four steps.
Since the validity assessment points out the rules which infer invalid solutions and the TURING Test experimentation reveals a so-called optimal solution to each test case, we are able to refine these rules with the objective to provide the optimal

(i.e. most dependable) solution. This, naturally, leads to an improved input-output behavior of the system, and thus to a more dependable system.

The benefit of this standardized validation framework is that developers of knowledge-based systems can reference it when describing the validation process to the end user. This may enhance the acceptability of the system.
Furthermore, this framework attempts to minimize the effort involved in validation of the expert system. This is because cases derived from the knowledge in the ***VKB*** don't have to be resolved in the process. The reason not to resolve them is that the ***VKB*** is intended to serve as a source of *external* knowledge, which consists of a historical solution that obtained good marks in the past.
Lastly, this minimized effort leads to reduced and more predictable costs.

A comprehensive description of all steps as well as the research behind this work can be found in (Knauf 2000). Also (Knauf et al. 2002) provides a more detailed description of this framework.

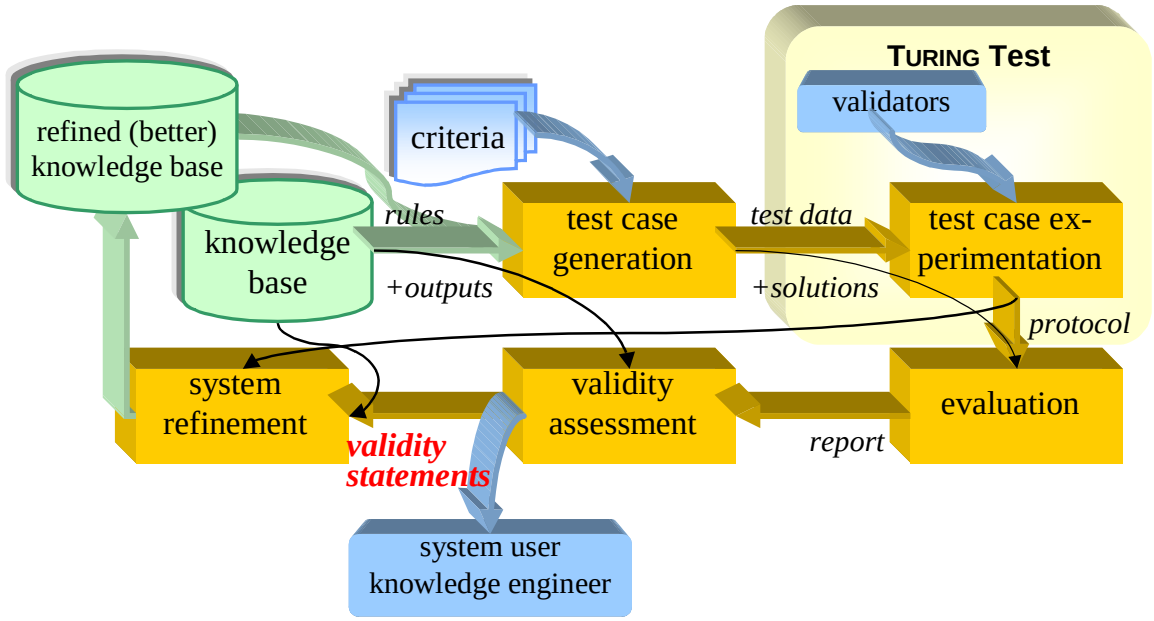

Figure 1: The Validation Framework of Knauf et al.

3. TSURUTA's Concept of VKB

In (Tsuruta et al. 2000a), a bi-directional, many-sided explanation typed multi-step validation method (*MMBV*) was proposed. Using this method, knowledge engineers and computers can share validation loads with experts. Thus, the validation load on busy experts is expected to decrease.

However, in order for knowledge engineers and computers to share more validation

load with experts, it is important for knowledge engineers and computers to share more validation knowledge with experts, through incorporating validation knowledge into computers. Thus, the concept of Validation Knowledge Base (***VKB***) and a validation approach based on ***VKB*** has been suggested (Tsuruta et al. 2000b, Tsuruta et al. 2002), which can reuse experts' validation experiences and has the effect of limiting the validation load on busy experts.
However, there is a serious problem called the "knowledge acquisition bottleneck". It seems even more difficult to acquire validation knowledge than to acquire domain knowledge because validation knowledge is a kind of meta-knowledge used for validating domain knowledge.

In order to solve this problem, the following approach was suggested in (Tsuruta et al. 2000b, Tsuruta et al. 2002). This approach is based on the concept, that computers, supported by knowledge engineers and experts, acquire, validate, and refine validation knowledge (***VKB***), based on the experts' validation data stored in the validation data base (***VDB***) of the validation system. An implementation detail of this concept is described in (Tsuruta et al. 2000a). Thus, correct and consistent validation knowledge can be easily acquired and incorporated as a Validation Knowledge Base (***VKB***), though such knowledge is difficult to acquire, also because experts are too busy to teach such validation expertise for various kinds of situations. Furthermore, this knowledge is often different or inconsistent depending on experts. This is explained more concretely as follows.

3.1 Experts' validation data base: VDB

In the above-mentioned ***VKB*** approach, the validation knowledge is acquired through experts' validation data in the ***VDB*** of the validation system such as the one introduced in (Tsuruta et al. 2000a).

VDB includes test cases. A test case consists of test data, and results (solutions) along with additional informal information as test process data. The test process data include the test schedule and delay status. Test results consist of solutions, explanations, validation results, and comments. Validation results include evaluators and evaluation values such as *OK* (valid), *NG* (invalid). Comments include any thoughts or explanations with a particular test case.

To summarize, ***VDB*** is structured as follows:

- test data: *test case input*
- test process data: *test schedule*, *delay status*
- test results:
  - solution
  - explanations
  - comments: *thoughts, explanations*
  - evaluator
  - evaluation result: *valid, invalid, degree of validity*

Thus, validation knowledge is automatically constructed and stored in the VKB as described below.

3.2 Validation knowledge base: VKB

As mentioned above, experts' validation data in the ***VDB*** includes test data (problems), solutions, and experts' validation results. They are considered to be experiences or examples of experts' validation knowledge. Therefore, the validation knowledge is acquired from the ***VDB*** and represented as a case library (Tsuruta et al. 2002) or as a rule-base (Tsurutra et al. 2000b). That is, Validation Knowledge Base (***VKB***) can be constructed from ***VDB*** by putting the test cases (problems with solutions), into a case-condition part (rule's condition part), and the experts' validation results (expert's evaluation value with comments) into a case-solution part (rule's conclusion part). Thus, the ***VKB*** contains the associated evaluation (validity degree) to each test case (i.e. to each pair [test data, solution]) along with additional knowledge about the circumstances of validation for each test case.

For example, as to a Traveling Salesman Problem, a case-condition part (rule's condition part) is a problem (test data) such as a list of visited cities and constraints, accompanied with its solution such as an optimally ordered sequence of visited cities. A case-solution part (rule's conclusion part) is the expert's evaluation value such as *OK* (valid), *NG* (invalid) or as grade of 1 to 5.

Each knowledge piece of the ***VKB*** has various properties, such as the confidence value (***CV***), many-sided explanation, expert's comment, etc. Further, in order to confirm the correctness of the acquired ***VKB***, it has also a property called ***Supporter***, which is the list of expert supporters who have accepted the knowledge piece, to trace back from where the validation knowledge originated (Tsuruta et al. 2002).

The validation and refinement of the acquired validation knowledge is necessary and important for correct validation. In the proposed approach, an acquired new validation knowledge piece (a new case or a new rule), for example, is checked in comparison with the existing ones in the ***VKB***. If an identical one is found, its confidence value (*CV*) is increased, and they are integrated into one knowledge piece. However, if inconsistency exists, the *CV* is decreased (Tsuruta et al. 2000b), and the experts' validation is retried to check validation knowledge by the responsible experts to be traced back. Other experts can be involved to assist if needed. That is, each piece of validation knowledge is validated and refined by the persons described in its ***Supporter*** property indicating the persons responsible for the knowledge, namely indicating experts who made or accepted the validation results (Tsuruta et al 2002). And, the wrong rule is removed or ignored under the control of *CV* or as a result of the above retrial.

Experts' validation knowledge can be easily acquired and incorporated as correct and consistent Validation Knowledge Base (***VKB***), though experts are too busy to teach or to validate such validation knowledge.

Thus, computers can automatically infer the validation results, utilizing the ***VKB***, and can further share the validation load of busy experts, with the help of knowledge engineers who check and modify the automatic validation results. As a result, the validation load of busy experts is lightened.

To summarize, the original VKB concept of Tsuruta is base on the following structured validation knowledge:

- test data
- solutions
- evaluators' validation results
  - an accepted solution
  - a confidence value: *a validity degree that changes according to new examples*
  - a many-sided explanation
  - experts' comments
  - a list of supporters: *experts, who support the solution*

4 TSURUTA's and UEHARA's Concept of VESA

The ***VESA*** concept adopts the idea of software agents in general and the recent developments in this field (Weiss 1997). In particular, SINGH and HUHNS (1997) address some basic concepts and assumptions as used here as well. However, advanced ideas like

- the issue of learning in general,
- the issue of cooperation and competition, and
- the issue of learning
  - about/from other agents and the world or
  - by communication and understanding

are far away from the fundamental agent concept introduced here.

With the view to future opportunities for temporarily replacing human input, the ***VKB*** itself is extended by a Validation Expert Software Agent (***VESA***) concept. This concept was originally introduced by TSURUTA and UEHARA. In the original concept ***VESA***s obtain and store validation knowledge / data autonomously from validation results of the experts participating in the test case experimentation.

This approach is adopted for the validation framework of KNAUF (Knauf et al. 2004a, Knauf et al. 2004b). A basic difference to the original concept is, that the "knowledge base" of a ***VESA*** does not really consist of validation knowledge. Instead, it is based knowledge about similarities in human expert responses to validation requests (solutions and ratings). This is a very different issue. The knowledge of a ***VESA*** can be compared with the (meta-) knowledge of a Case Based Reasoning (***CBR***) system's "inference engine", which

a) looks for a "most similar case" in a library of historical cases,
b) adopts its solution towards a solution of the present case and
c) maintains itself by adding new (or removing old) cases while ensuring consistency.

Of course, the maximum similarity can be reached, when a ***VESA*** simply adopts a most recent former response of its human origin to the same request. Typically, such a response is not available and similarities with other human expert responses must be analyzed.

The basic assumption of the new ***VESA*** concept is that experts who provide similar solutions to test cases and similar ratings to other experts' solutions might have a similar knowledge structure. Therefore, an expert might be modeled by an agent that provides the response of another human expert, who had a maximum similarity with the considered expert in the past. Since knowledge structures do change over time

the degree of similarity depends on both the ratio of same reflections (solutions, ratings) and the „age“ of this identical behavior.

VESA can only be used, if there is enough historical knowledge to derive a “most similar (human) expert”. Even after a learning period a long term use of a ***VESA*** instead of its human origin is useless, because the derived similarities change over time, so that the base to perform ***VESA*** becomes obsolete if it is not “maintained” by employing their human origins from time to time.

In fact, the modeling of individual human expertise by ***VESA*** is performed just like in a ***CBR*** system. This rises the question:

> *If **CBR** is really the right approach to validate a system’s performance – why don’t we use this approach for the system itself?*

This question has several answers:

1. The objective of ***VESA*** is not a better quality of validation knowledge. In fact, the opposite is true: If we substitute an original source of knowledge by its model, there is a risk, that it gets worse. The real objective is to decrease the human involvement and thus, the validation expenses. Finally, it is a trade-off between the desired quality of validation results and the costs to reach them.
2. ***VESA*** does not really model validation knowledge; this is done in a much more sophisticated way by ***VKB***. Instead, ***VESA*** models a very individual behavior of a particular human expert while performing a validation task by providing a most similar one. This is done by analyzing similarities of the (original) validation knowledge in-between particular former members of expert panels. ***VESA*** itself contains knowledge about validation behavior (which is somehow “meta – validation knowledge”), but not validation knowledge itself. For this meta-knowledge the ***CBR*** approach seems to be the right one. This must not necessarily be the case for the topical knowledge of the application field.
3. ***CBR*** systems usually have an adaptation function, which computes a solution to a present case by adapting the solution of the most similar case. Here, this “adaptation” is just consists in simply providing the same solution. The ***VESA*** approach so far might have has some potentials for further development and refinement by considering ***CBR*** adaptation techniques.

One might also argue, that

- on the one hand the TURING Test technique requires the use of high quality human expertise but,
- on the other hand, ***VESA***s knowledge is (of course) behind the human standard of quality.

This is very true and we have to carefully validate the concept and to specify a tolerable loss of quality when using a ***VESA***.

Indeed, this issue led to a controversial discussion between the authors of this report. Finally, we agreed to have a look at the experimentation results:

- In case these results reveal the usefulness of VESA for substituting its human origin, it’s worth to be developed further.
- Otherwise this approach needs to be revised.

Thus, the experimentation reported here aims at the validation of this concept.

Each ***VESA*** is an autonomous software agent corresponding to a particular human expert. It gains personal validation knowledge mainly from personal data such as (not always best) solutions, ratings, etc. of the human expert validator corresponding to it. Furthermore, is can be considered to be a model that represents the validation experience and behavior of a group or an organization of validation experts.

In every validation session, the ***VESA***s become more intelligent as well as more adaptive to wider (similar but slightly different) applications, since they can learn from test inputs, the associated answers, their certainties and their ratings provided by the human validators. Namely, they increase their validation competence through validation knowledge gained by various sessions over time.

Though a ***VESA*** is a model of a human validation expert, it can also gain the validation knowledge / data of other validators, when a very high-rated (but not always best) solution happens to be derived by one of the same type of validators which usually have almost the same solutions. Since they are not human but machine, anonymity will be kept even if they get information from other (human) experts. They do not need the name of each expert, but rather an *ID* to distinguish whether or not the information belongs to the same expert.

5 Incorporating VKB and VESA in the Validation Framework

5.1 VKB

5.1.1 Content of VKB

The information that needs to be stored and maintained in the ***VKB*** for use in the ***test case experimentation***, consists of the required input data, the produced output data, and some additional necessary data. According to the formal settings in (Knauf et al. 2002) and (Kurbad 2003), the ***VKB*** contains a set of historical test cases, which can be described by 8-tuples

$$[\, t_j\,,\, E_K\,,\, E_I\,,\, sol_{Kj}^{opt}\,,\, r_{IjK}\,,\, c_{IjK}\,,\, \tau_S\,,\, D_C\,]$$

with

- t_j being a test data (a test case input),
- sol_{Kj}^{opt} being a solution associated to t_j ,
- E_K being a list of experts who provided this particular solution,
- E_I being a list of experts who rated this solution,
- r_{IjK} being the rating of this solution, which is provided by the experts in E_I,
- c_{IjK} being the certainty[2] of this rating,
- τ_S being a time stamp associated with the validation session in which the rating was provided, and
- D_C being an informal description of the application domain ***C*** that is helpful to explain similarities between different domains or fields of knowledge.

[2] Besides providing a rating that might be ***0*** (wrong) or ***1*** (correct), the experts have the opportunity to express, whether (***c=1***) or not (***c=0***) they feel certain while providing this rating.

Additionally, a ***list of supporters*** $E_S \subseteq E_I$ for each solution sol_{Kj}^{opt} is derived from this data. E_S is the list of rating experts, who provided a positive rating for sol_{Kj}^{opt}.

The **VKB** is not completely transparent to all agents in the validation process. According to the purpose of use, some of the data is hidden to certain agents. For example, to ensure the anonymity while solving and rating test cases within the TURING Test, E_K and E_I must not be presented to the expert panel of the current session. Furthermore, to ensure an unbiased rating, the historical rating r_{IjK} must not be presented to the expert panel that currently rates the solution.

5.1.2 Usage of VKB

5.1.2.1 Involvement of the VBK in the Test Case Experimentation

The intermediate results that occur during the experimentation as well as the ***VKB*** itself are stored in a relational database by using a ***client-server database management system*** (**DBMS**), which provides decentralized access to centralized data for clients, which work independently from each other. All data is kept central to the view of ***knowledge engineer*** (server), while only the necessary parts of it are shown to the ***expert panel*** (client) (Kurbad 2003).

All experts of the panel take part in the experimentation session independently. By utilizing an HTML-based implementation approach for the client application, each expert is almost free in the choice of time and place of his work. This effectively limits delays that are caused by experts who would otherwise be unavailable as well as the costs of the whole validation process.

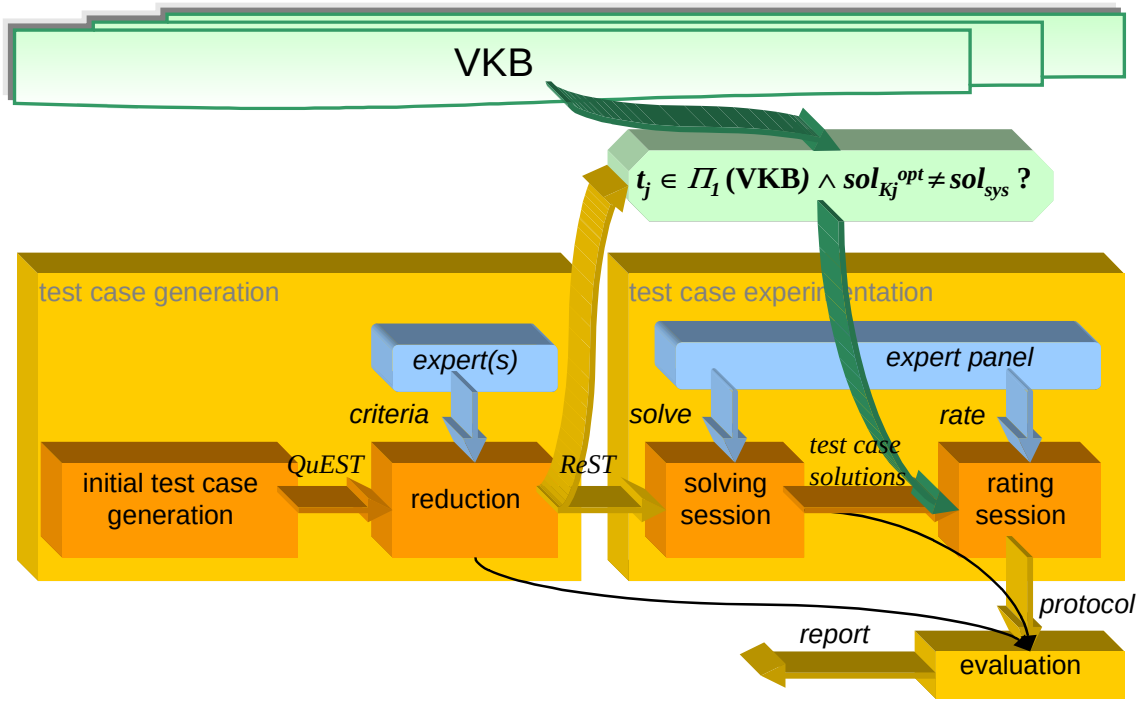

Figure 2: Involvement of the VBK in the Test Case Experimentation

Figure 2 sketches the usage of ***VKB*** in the test case experimentation. After generating the so-called ***Quasi Exhaustive Set of Test Cases*** (***QuEST***) (Knauf et al. 2002) both ***QuEST*** and the historical cases in ***VKB*** are a subject of the criteria-based reduction procedure which aims at a subset of cases in ***QuEST*** or ***VKB***, which meets the requirements of the current application and is small enough to be the subject of the test case experimentation.

Since a ***VKB*** is a database of test cases and their associated solutions, which received an optimal rating in previous validation sessions, these solutions are considered as an additional (external) source of expertise that does not explicitly appear in the solving session. Therefore, the cases originates from the ***VKB*** are not a subject of the test case solving session.

Regardless of their former ratings, the cases from the ***VKB*** have to be rated by the current expert panel again for two reasons:

1. Topical domain knowledge of AI systems has a dynamic nature. It might have changed since the time, when the information in the ***VKB*** has been acquired. This might be due to recent insights, but also in changed application circumstances, for example.
2. Additionally, there is a responsibility for the results of applying the validation technology, i.e. for the validity statements as well as for the refined knowledge base. These results need, when communicated and used for (commercial, political, …) decisions, a clear association to responsible persons. Of course, the current panel which rated the solutions must serve as these responsible persons.

Albeit there is already a (historical) rating for the test cases in the ***VKB***, this panel must have the opportunity to provide own ratings to these test cases.[3]

Fortunately, not all cases of the ***VKB*** that "survived" the criteria-based reduction process need to be rated again: Only cases with solutions different from the system's solution have to be involved in the rating process (see ***sol= system's ?*** - box in figure 1), because

(1) we are only interested in new external knowledge that is outside the expertise of the expert panel and

(2) the system's solution is in the process anyway[4] and the test case solving procedure additionally provides alternative ("man-made") solutions to it.

5.1.2.2 Utilizing the Experience of VKB

As previously indicated, the knowledge gained in the ***VKB*** is also applied for other useful purposes:

1. It can be used for a refined competence estimation of the panel experts. This estimation is used as a weight an expert's rating of the system's solution to compute its validity degree (Knauf et al. 2002). Since all resulting validity statements are derived from these validity degrees, the refinement of the competence estimation leads to improved results of the entire technology. In fact, the consequence of better validity statements is a "more dependable" system after its refinement. Furthermore, this competence estimation is useful to select appropriate experts for the panel of upcoming sessions.
2. Second, the ***VKB*** can support the identification of the optimal solution, which is the basis for the system refinement and the updating process of the ***VKB*** itself. If several solutions are candidates to be the "optimal solution", i.e. they receive the same approval by the expert panel, the information in the ***VKB*** is helpful to differentiate these candidates.

Both approaches are introduced in (Kurbad 2003) and firstly sketched in the international public here.

5.1.2.2.1 Competence Estimation of the Experts

Since the competence estimation of the experts is based on the experts' performance in the rating session, the ratings and their certainties for the test cases originated from the ***VKB*** needs to be included in the estimation.

Since the ***VKB*** holds knowledge about the experts' competence in previous sessions, i.e. "historical competence", it opens the chance to select an appropriate expert panel for a scheduled session. Derived from the information in the ***VKB***, we formally introduced

1. a so-called **historical session competence** $sess\_est_{hist}(e_i, S'_i)$ of a certain expert e_i within a session S'_i,

[3] Nobody would agree to be responsible for something that he/she can not control.

[4] The test case generation step exclusively produces test cases with system's solutions.

2. a **historical competence trend** $trnd\_est_{hist}(e_i)$, which describes the development of an expert's e_i competence over time,
3. a **competence gain** $\Delta sess\_est_{hist}(e_i, \sigma^t_i)$ from one session to the next and an **average competence gain** $\delta_i(\sigma^t_i)$ over time,
4. a classification of experts as those with an (a) *increasing*, (b) *even*, and (c) *decreasing* competence over time, and
5. an **average historical competence**} $avg\_est_{hist}(e_i)$.

Finally, we developed a guideline to use these concepts for the selection of an appropriate expert panel.[5]

5.1.2.2.2 Identification of the Optimal Solution

For the final **system refinement** step of the entire validation technology, the concept of an optimal solution was introduced in (Knauf et al. 2002). This is, loosely speaking, the solution $sol_{opt}(t_j)$ to a test data t_j that gained the maximum approval by the experts in the current panel.

Unfortunately, it might happen that there are several solutions, which enjoy the maximum approval. In these cases, the ***VKB*** is used to identify one of them as the "very best" one.

For this purpose, we introduced a step-by-step filtering process that is applied until one candidate solution is left over:

1. Firstly, the average competence of the experts, who are in the ***VKB***'s **list of supporters** of the candidate solutions are considered. The candidate solution, which enjoys the maximal support by the ***VKB***, is considered the "very best" one.
2. In case there are still several solutions as the outcome of the step above, a **list of vetoers**[6] is derived from the ***VKB*** and their average competence is calculated by using the ***VKB***. The candidate solution, which received the minimal "resistance" by the ***VKB***, is considered the "very best" one.
3. If there are still several candidate solutions after these two steps, the supporters for each of the remaining candidate solutions are compared: The solution that is supported by the expert e_i with the maximal competence $cpt(e_i,t_j)$ for the test data t_j, is considered the "very best" one.
4. The last chance to identify the "very best" solution, in case there are still several ones after these three steps, we perform a "run-off" session with the expert panel and the remaining candidate solutions.

5.2 VESA

5.2.1 Sources of VESA's Knowledge

[5] Note, that the authors themselves claim to utilize these estimations with care, because they are based on data, which might be incomplete, irrelevant, and not representative. Furthermore, social reasons require to handle all the concepts about an expert's competence with care, discretion and social responsibility.

[6] Vetoers are experts, who provided a negative rating for a considered solution.

The knowledge base to dynamically form a ***VESA*** in case of its need is simple: Gaining all information that is available. For each human expert it keeps

(1) each and every solution he/she provided to a test data,
(2) each and every rating he/she provided to a solution in
(3) each and every historic session indicated by a time stamp.

5.2.2 Dynamic Construction of VESA

Providing a solution by *VESA*

In case e_i solved (with a solution different from „unknown“) or rated t_j in former sessions, his/her provided or as (with certainty 1) „correct“ rated solution with the latest time stamp τ_S will be provided by ***VESA***.

If e_i never considered a case like t_j before, similarities with other experts (which might have the same "school" or "thinking structures") are considered. Among all experts, who ever delivered a solution to t_j , the one with the largest subset of the solutions or ratings like e_i -s to the other cases is identified as the expert e_{sim} with the most similar behavior.

$$Solver_i^0 := \{e' : [t_j, E_K, E_I ...] \in VKB \wedge (e' \in E_K \vee e' \in E_I)\}$$

$$Solver_i^1 := e_{sim} : (e_{sim} \in Solver_i^0) \wedge (|\{[[t_j, e_i, sol_{ij}, \tau_s], [t_j, e_{sim}, sol_{sim,j}, \tau_s]] : sol_{ij} = sol_{sim,j}\}| \rightarrow \max!)$$

His/her latest provided or as “correct” rated solution is assumed to be the one of e_i as well, and thus provided by ***VESA***.
If there is no such most similar expert, the solution *unknown* will be provided by ***VESA***.

Providing a rating by VESA

In case a ***VESA*** is requested to provide a rating to a given solution, similar considerations lead to an "assumed rating" of e_i:

1. If e_i considered (solved or rated) the same test case t_j in former sessions, we look at the rating or the provided solution with the latest time stamp:
    a. In case the latest consideration is a rating ***r*** along with a certainty ***c***, both the same rating ***r*** and the same certainty ***c*** are adopted and provided by ***VESA***.
    b. In case the latest consideration is a provided solution ***sol*** (different from *unknown*), ***VESA*** provides for this solution a rating ***r:=1*** (correct) and a certainty ***c:=1*** (for sure) and for all other solutions a rating ***r:=0*** (incorrect) and a certainty ***c:=1***.
2. If e_i never considered (solved or rated) the test case t_j in former sessions, we look for a "most similar" expert e_{sim} who solved this case, i.e. a one who

provided the largest amount of the same solutions and/or ratings in the same session to other cases in the past:

$$Solver_i^0 := \{e' : ([t_j, E_K, E_I ...] \in VKB) \wedge ((e' \in E_K) \vee (e' \in E_I))\}$$
$$Solver_i^1 := e_{sim} : (e_{sim} \in Solver_i^0) \wedge$$
$$(|\{[[t_j, e_i, sol_{ij}, \tau_s], [t_j, e_{sim}, sol_{sim,j}, \tau_s]] : sol_{ij} = sol_{sim,j}\} \cup$$
$$\{[[t_j, e_k, e_i, sol_{kj}, r_{ijk}, \tau_s], [t_j, e_k, e_{sim}, sol_{kj}, r_{sim,j,k}, \tau_s]] : r_{ijk} = r_{sim,j,k}\} |\rightarrow \max!)$$

a. If the latest consideration of t_j by e_{sim} is a rating ***r*** along with a certainty ***c***, ***VESA*** adopts and provides both.
b. If the latest consideration of t_j by e_{sim} is a solution ***sol***, ***VESA*** provides for this solution a rating ***r:=1*** (correct) and a certainty ***c:=1*** (for sure) and for all other solutions a rating ***r:=0*** (incorrect) and a certainty ***c:=1***.
c. If there is no „most similar“ expert e_{sim}, provide the rating ***r := norating*** and a certainty ***c := 0*** .

As a future benefit of the ***VESA***s we expect that

1. ***VESA*** can replace the human expert when he/she is too busy or too expensive to participate in validation,
2. ***VESA*** can be a competent validator and upgrade the test case experimentation and
3. a group of ***VESA***s might do test case experimentation without experts, since they have different validation knowledge and can be tested from various views.

Therefore, the ***VESA*** concept brings a really new dimension into the validation technology by displacing human input systematically to software agents.

Actually, to learn a model of the human experts' problem solving behavior, ***VESA*** still depends on the knowledge of human validators. Learning in the concept of ***VESA*** is analyzing the solving and rating behavior of human their origins. The quality of the learning results, i.e. the quality of ***VESA***, depends on the quantity and coverage of data provided by the human experts. Therefore,

- on the one hand, a ***VESA*** is able to replace its human origin temporary, but,
- on the other hand, a ***VESA*** becomes worse in case of missing human input over a long period.

6 The Prototype Application

6.1 The Knowledge Base

By consulting the topical literature, the first author derived the following informal knowledge.

6.1.1 Initial Informal Knowledge

General Rules

- Meals that are rich in content call for a wine that is rich in body.

- Light meals call for a light wine.
- Premium meals call for a fine and premium wine.

Particular Rules
Meat
- Light colored meat, such as fowl and veal call for a fruity, grapy red wine with less tannin.
- Fried and grilled meats call for a young red wine rich in tannin.
- With smoked meat there is a correlation between the length of time in the meal's preparation and the time to mature the wine. Furthermore, tannins help to make the food digestible. Thus, a mature Barolo or a mature Brunello fits well.
- The autumnal and slightly sweet taste of venison calls for a strong partner. A dark, fruit-accentuated red wine from the "new world" is appropriate. Alternatively, a mature red wine from Burgundy, Bordeaux or the Rhone-Valley is acceptable.

Fish
- Steamed fish calls for a light, fresh and low acid-accentuated white wine. An alternative is a dry, fruity, low tannin Rosé.
- Fried or grilled Fish has an intensive taste and gets along well with a (possibly in a wooden barrel matured) white wine or a red wine that has not too much tannin. To summarize, a strong white wine or a low tannin red wine is acceptable.

Asian meals
- The intensive flavoring and spice fits with the freshness and intensity of an aromatic white wine. Muscatel, Gewürztraminer, Sauvignon Blanc, or a semi-dry Riesling are appropriate wines.

Cheese
- Hard cheese calls for a white wine that is rich in content. Or a velvet, low tannin red wine, especially Pinot Noir or Amarone.
- Soft cheese needs a similar wine that hard cheese, but a little lighter. Beaujolais is also acceptable.
- Goat cheese calls for a dry and fruity white wine.
- Blue mold cheese fits well with any wine other than sweet ones.

Desserts
The switch in taste that comes with the dessert needs a switch in the wine taste as well.
- Fruit dessert fits well with Riesling that is rich in acid.
- Aromatic desserts (flavored with cloves, anise, or cinnamon, e.g.) call for a Gewürztraminer.
- Ice cream fits best with Prot Wine.

This informal knowledge reveals that in particular for Asian cooking style, the rules are quite general. Since the experiment involves Asian experts as well, it is expected that this part of the knowledge base will be improved exceptionally.

6.1.2 Towards Formalizing the Knowledge

The first step towards formalizing this knowledge is to define input variables along with their value rage as well as the outputs that need to be distinguished:

1. the main ingredient,
2. the kind of preparation, and
3. the style of preparation.

Thus, the input space is spanned by three input variables with the following value sets:

I = { [s_1, s_2, s_3] :
s_1 ∈ {pork, beef, veal, venison, fowl, meat, fish, hard cheese, soft cheese, goat cheese, blue mold cheese, fruit dessert, aromatic dessert, ice cream}
s_2 ∈ {non (raw), steamed, boiled, grilled, fried, stewed, casserole, deep fried}
s_3 ∈ {Asian, Western } **}**

The theoretical number input combinations 14 x 8 x 2 = 224. In fact, some of them don‘t make sense: grilled ice cream, e.g.

Output space: 24 different wine grades $O = \{ o_1, o_2, \ldots, o_{24} \}$
o_1: Red wine, fruity, low tannin, less compound
o_2: Red wine, young, rich of tannin
o_3: Red wine, dark, fruity, from the „new world“
o_4: Red wine, mature, from the Rhone valley (France)
o_5: Red wine, velvet, low tannin
o_6: Pinot Noir
o_7: Amarone
o_8: Burgundy, mature
o_9: Bordeaux, mature
o_{10}: Barolo, mature
o_{11}: Brunello, mature
o_{12}: Beaujolais
o_{13}: Rosé, dry, fruity, low tannin
o_{14}: White wine, light, fresh, low acid
o_{15}: White wine, strong, low tannin
o_{16}: White wine, rich in content
o_{17}: White wine, dry, fruity
o_{18}: Muscatel
o_{19}: Gewürztraminer
o_{20}: Sauvignon Blanc
o_{21}: Riesling, semi dry
o_{22}: Riesling, rich of acid
o_{23}: Port wine
o_{24}: Any wine besides smooth one

Expressing the informal knowledge with these input and output specification with HORN clauses leads to the following rule base.

6.1.3 Formal Knowledge Base

The Rule base R consists of the following rule set $R = \{ r_1, r_2, \ldots, r_{38} \}$ in which rules with the same input variables (but different values) and the same output have a second (sub-) index:

r_1: Red wine, fruity, low tannin, less compound ← (main ingredient = fowl)

r_2: Red wine, fruity, low tannin, less compound ← (main ingredient = veal)

r_3: Red wine, young, rich of tannin ← (main ingredient = pork) ∧ (preparation = grilled)

r_4: Red wine, young, rich of tannin ← (main ingredient = pork) ∧ (preparation = fried)

r_5: Red wine, young, rich of tannin ← (main ingredient = beef) ∧ (preparation = grilled)

r_6: Red wine, young, rich of tannin ← (main ingredient = beef) ∧ (preparation = fried)

r_7: Red wine, fruity, low tannin, less compound ← (main ingredient = fowl)

r_8: Red wine, fruity, low tannin, less compound ← (main ingredient = veal)

$r_{9.1}$: Barolo, mature ← (main ingredient = pork) ∧ (preparation = stewed)

$r_{9.2}$: Barolo, mature ← (main ingredient = beef) ∧ (preparation = stewed)

$r_{9.3}$: Barolo, mature ← (main ingredient = veal) ∧ (preparation = stewed)

$r_{9.4}$: Barolo, mature ← (main ingredient = venison) ∧ (preparation = stewed)

$r_{9.5}$: Barolo, mature ← (main ingredient = fowl)

$r_{10.1}$: Brunello, mature ← (main ingredient = pork) ∧ (preparation = stewed)

$r_{10.2}$: Brunello, mature ← (main ingredient = beef) ∧ (preparation = stewed)

$r_{10.3}$: Brunello, mature ← (main ingredient = veal) ∧ (preparation = stewed)

$r_{10.4}$: Brunello, mature ← (main ingredient = venison) ∧ (preparation = stewed)

$r_{10.5}$: Brunello, mature ← (main ingredient = fowl) ∧ (preparation = stewed)

r_{11}: Red wine, dark, fruity, from the „new world“ ← (main ingredient = venison)

r_{12}: Burgundy, mature ← (main ingredient = venison)

r_{13}: Bordeaux, mature ← (main ingredient = venison)

r_{14}: Red wine, mature, from the Rhone valley ← (main ingredient = venison)

r_{15}: White wine, light, fresh, low acid ← (main ingredient = fish) ∧ (preparation = steamed)

r_{16}: Rosé, dry, fruity, low tannin ← (main ingredient = fish) ∧ (preparation = steamed)

r_{17}: White wine, strong, low tannin ← (main ingredient = fish) ∧ (preparation = fried)

r_{18}: White wine, strong, low tannin ← (main ingredient = fish) ∧ (preparation = grilled)

r_{19}: Red wine, fruity, low tannin, less compound ← (main ingredient = fish) ∧ (preparation = fried)
r_{20}: Red wine, fruity, low tannin, less compound ← (main ingredient = fish) ∧ (preparation = grilled)
r_{21}: Muscatel ← (style = Asian)
r_{22}: Gewürztraminer ← (style = Asian)
r_{23}: Sauvignon Blanc ← (style = Asian)
r_{24}: Riesling, semi dry ← (style = Asian)
r_{25}: White wine, rich in content ← (main ingredient = hard cheese)
r_{26}: Red wine, velvet, low tannin ← (main ingredient = hard cheese)
r_{27}: Pinot Noir ← (main ingredient = hard cheese)
r_{28}: Amarone ← (main ingredient = hard cheese)
r_{29}: White wine, rich in content ← (main ingredient = soft cheese)
r_{30}: Red wine, velvet, low tannin ← (main ingredient = soft cheese)
r_{31}: Pinot Noir ← (main ingredient = soft cheese)
r_{32}: Amarone ← (main ingredient = soft cheese)
r_{33}: Beaujolais ← (main ingredient = soft cheese)
r_{34}: White wine, dry, fruity ← (main ingredient = goat cheese)
r_{35}: Any wine besides smooth one ← (main ingredient = blue mold cheese)
r_{36}: Riesling, rich of acid ← (main ingredient = fruit dessert)
r_{37}: Gewürztraminer ← (main ingredient = aromatic dessert)
r_{38}: Port wine ← (main ingredient = ice cream)

6.2 Initial Test Cases

According to (Knauf et al. 2002) and (Knauf 2000), test case generation is divided in the following steps and sub steps:

1. **Generation of a Quasi Exhaustive set of test Cases (*QuEST*)**
    1. **Computation of dependency sets** In this step is analyzed how the elements of ***O*** (the system's outputs) depend on the elements of the rule set ***R*** as well as the elements of the sensor data set ***S***.
    2. **Computation of critical values and scanning distances** Here, the region boundary values (see above) and a practically useful scanning

distance for numerical input data will be determined. A critical value is a boundary value; a scanning distance is the distance of a (future) test data from a boundary of a numerical input dimension.

3. **Computation of the sets of potential test case values** This step is to determine values s_i' for each sensor data $s_i \in S$ that are meaningful for the composition of ***QuEST***. This is done for each output $o_i \in O$ separately, because not every sensor contributes to a particular output. This limits the cardinality of ***QuEST***.
4. **Composing the set of all potential test data** Here, the test data $t_j = [s_1^j, s_2^j, \ldots, s_m^j] \in I$ is composed based on the potential test case values.
5. **Minimizing the set of all potential test data** In the last step, the test data set computed so far will be minimized by excluding test data which are subsumed by others. These are the test data which are located in between two others of the same region of influence in the input space.

2. **Generation of a Reasonable Set of test Cases (*ReST* $\subseteq$ *QuEST*)** in case ***QuEST*** is too large to be considered by the human experts by using application dependent criteria

6.2.1 Generation of a Quasi Exhaustive Set of Test Cases (*QuEST*)

6.2.1.1 Computation of dependency sets

Here, we compute for each $o_i \in O$ a rule dependency set $R_i \subseteq R$ that contains the rules $r_k \in R$ on which o_i depends and a sensor dependency set $S_i \subseteq S$ that contains sensor variables $s_k \in S$ on which o_i depends. The result is as follows:

$R_1 = \{ r_1, r_2, r_7, r_8, r_{19}, r_{20} \}$	$R_{13} = \{ r_{16} \}$	$S_1 = \{ s_1, s_2 \}$	$S_{13} = \{ s_1 , s_2 \}$
$R_2 = \{ r_3, r_4, r_5, r_6 \}$	$R_{14} = \{ r_{15} \}$	$S_2 = \{ s_1, s_2 \}$	$S_{14} = \{ s_1 , s_2 \}$
$R_3 = \{ r_{11} \}$	$R_{15} = \{ r_{17}, r_{18} \}$	$S_3 = \{ s_1 \}$	$S_{15} = \{ s_1 , s_2 \}$
$R_4 = \{ r_{14} \}$	$R_{16} = \{ r_{25}, r_{29} \}$	$S_4 = \{ s_1 \}$	$S_{16} = \{ s_1 \}$
$R_5 = \{ r_{26} , r_{30} \}$	$R_{17} = \{ r_{34} \}$	$S_5 = \{ s_1 \}$	$S_{17} = \{ s_1 \}$
$R_6 = \{ r_{27}, r_{31} \}$	$R_{18} = \{ r_{21} \}$	$S_6 = \{ s_1 \}$	$S_{18} = \{ s_3 \}$
$R_7 = \{ r_{28}, r_{32} \}$	$R_{19} = \{ r_{22} , r_{37} \}$	$S_7 = \{ s_1 \}$	$S_{19} = \{ s_1, s_3 \}$
$R_8 = \{ r_{12} \}$	$R_{20} = \{ r_{23} \}$	$S_8 = \{ s_1 \}$	$S_{20} = \{ s_3 \}$
$R_9 = \{ r_{13} \}$	$R_{21} = \{ r_{24} \}$	$S_9 = \{ s_1 \}$	$S_{21} = \{ s_3 \}$
$R_{10} = \{ r_9 \}$	$R_{22} = \{ r_{36} \}$	$S_{10} = \{ s_1, s_2 \}$	$S_{22} = \{ s_1 \}$
$R_{11} = \{ r_{10} \}$	$R_{23} = \{ r_{38} \}$	$S_{11} = \{ s_1, s_2 \}$	$S_{23} = \{ s_1 \}$
$R_{12} = \{ r_{33} \}$	$R_{24} = \{ r_{35} \}$	$S_{12} = \{ s_1 \}$	$S_{24} = \{ s_1 \}$

6.2.1.2 Computation of critical values and scanning distances

Since there is no numerical input sensor, and no input has a reasonable ordering relation in-between its possible values, all values of s_1, s_2, and s_3 are considered critical.

S_1^{krit} = {pork, beef, veal, venison, fowl, fish, hard cheese,
soft cheese, goat cheese, blue mold cheese, fruit dessert,

aromatic dessert, ice cream}
S_2^{krit} = {non (raw), steamed, boiled, grilled, fried, stewed, casserole, deep fried}
S_3^{krit} = {Asian, Western}

Scanning distances don't apply in this case.

6.2.1.3 Computation of the sets of potential test case values

$V_{i,j}$ is a set of potential values of the sensor (input) variable s_j to validate the output o_i. According to (Knauf et al. 2002) and (Knauf 2000) this leads to the following $V_{i,j}$. The original approach suggests a "normal value" as the only element of a V_{ij} , if o_i does not depend on s_j. Since in this application domain there is no "normal value", we use the "Element" ***any*** for now.

$V_{1,1} = S_1^{krit}$	$V_{1,2} = S_2^{krit}$	$V_{1,1} = \{ any \}$
$V_{2,1} = S_1^{krit}$	$V_{2,2} = S_2^{krit}$	$V_{2,3} = \{ any \}$
$V_{3,1} = S_1^{krit}$	$V_{3,2} = \{ any \}$	$V_{3,3} = \{ any \}$
$V_{4,1} = S_1^{krit}$	$V_{4,2} = \{ any \}$	$V_{4,3} = \{ any \}$
$V_{5,1} = S_1^{krit}$	$V_{5,2} = \{ any \}$	$V_{5,3} = \{ any \}$
$V_{6,1} = S_1^{krit}$	$V_{6,2} = \{ any \}$	$V_{6,3} = \{ any \}$
$V_{7,1} = S_1^{krit}$	$V_{7,2} = \{ any \}$	$V_{7,3} = \{ any \}$
$V_{8,1} = S_1^{krit}$	$V_{8,2} = \{ any \}$	$V_{8,3} = \{ any \}$
$V_{9,1} = S_1^{krit}$	$V_{9,2} = \{ any \}$	$V_{9,3} = \{ any \}$
$V_{10,1} = S_1^{krit}$	$V_{10,2} = S_2^{krit}$	$V_{10,3} = \{ any \}$
$V_{11,1} = S_1^{krit}$	$V_{11,2} = S_2^{krit}$	$V_{11,3} = \{ any \}$
$V_{12,1} = S_1^{krit}$	$V_{12,2} = \{ any \}$	$V_{12,3} = \{ any \}$
$V_{13,1} = S_1^{krit}$	$V_{13,2} = S_2^{krit}$	$V_{13,3} = \{ any \}$
$V_{14,1} = S_1^{krit}$	$V_{14,2} = S_2^{krit}$	$V_{14,3} = \{ any \}$
$V_{15,1} = S_1^{krit}$	$V_{15,2} = S_2^{krit}$	$V_{15,3} = \{ any \}$
$V_{16,1} = S_1^{krit}$	$V_{16,2} = \{ any \}$	$V_{16,3} = \{ any \}$
$V_{17,1} = S_1^{krit}$	$V_{17,2} = \{ any \}$	$V_{17,3} = \{ any \}$
$V_{18,1} = \{ any \}$	$V_{18,2} = \{ any \}$	$V_{18,3} = S_3^{krit}$
$V_{19,1} = S_1^{krit}$	$V_{19,2} = \{ any \}$	$V_{19,3} = S_3^{krit}$
$V_{20,1} = \{ any \}$	$V_{20,2} = \{ any \}$	$V_{20,3} = S_3^{krit}$
$V_{21,1} = \{ any \}$	$V_{21,2} = \{ any \}$	$V_{21,3} = S_3^{krit}$
$V_{22,1} = S_1^{krit}$	$V_{22,2} = \{ any \}$	$V_{22,3} = \{ any \}$
$V_{23,1} = S_1^{krit}$	$V_{23,2} = \{ any \}$	$V_{23,3} = \{ any \}$
$V_{24,1} = S_1^{krit}$	$V_{24,2} = \{ any \}$	$V_{24,3} = \{ any \}$

6.2.1.4 Composing the set of all potential test data

According to (Knauf et al. 2002) and (Knauf 2000), these sets need to be composed into a potential test data set ***P*** by

$$P = \bigcup_{i=1}^{24} \prod_{j=1}^{3} V_{i,j}$$

With respect to (removable) repetitive elements, this leads to

$$P = (S_1^{krit} \times S_2^{krit} \times \{any\}) \cup (S_1^{krit} \times \{any\} \times \{any\}) \cup (\{any\} \times \{any\} \times S_3^{krit}) \cup (S_1^{krit} \times \{any\} \times S_3^{krit})$$

With respect to the cardinalities $|S_1^{krit}|$ = 13, $|S_2^{krit}|$ = 8, $|S_3^{krit}|$ = 2, and $|\{any\}|$ = 1 this leads to 13∗8∗1 + 13∗1∗1 + 1∗1∗2 + 13∗1∗2 = 145 potential test data $p_1, \ldots, p_{145}$:

$(S_1^{krit} \times S_2^{krit} \times \{any\})$:

t_1	pork	non (raw)	any	t_{53}	hard cheese	fried	any
t_2	pork	steamed	any	t_{54}	hard cheese	stewed	any
t_3	pork	boiled	any	t_{55}	hard cheese	casserole	any
t_4	pork	grilled	any	t_{56}	hard cheese	deep fried	any
t_5	pork	fried	any	t_{57}	soft cheese	non (raw)	any
t_6	pork	stewed	any	t_{58}	soft cheese	steamed	any
t_7	pork	casserole	any	t_{59}	soft cheese	boiled	any
t_8	pork	deep fried	any	t_{60}	soft cheese	grilled	any
t_9	beef	non (raw)	any	t_{61}	soft cheese	fried	any
t_{10}	beef	steamed	any	t_{62}	soft cheese	stewed	any
t_{11}	beef	boiled	any	t_{63}	soft cheese	casserole	any
t_{12}	beef	grilled	any	t_{64}	soft cheese	deep fried	any
t_{13}	beef	fried	any	t_{65}	goat cheese	non (raw)	any
t_{14}	beef	stewed	any	t_{66}	goat cheese	steamed	any
t_{15}	beef	casserole	any	t_{67}	goat cheese	boiled	any
t_{16}	beef	deep fried	any	t_{68}	goat cheese	grilled	any
t_{17}	veal	non (raw)	any	t_{69}	goat cheese	fried	any
t_{18}	veal	steamed	any	t_{70}	goat cheese	stewed	any
t_{19}	veal	boiled	any	t_{71}	goat cheese	casserole	any
t_{20}	veal	grilled	any	t_{72}	goat cheese	deep fried	any
t_{21}	veal	fried	any	t_{73}	blue mold cheese	non (raw)	any
t_{22}	veal	stewed	any	t_{74}	blue mold cheese	steamed	any
t_{23}	veal	casserole	any	t_{75}	blue mold cheese	boiled	any
t_{24}	veal	deep fried	any	t_{76}	blue mold cheese	grilled	any
t_{25}	venison	non (raw)	any	t_{77}	blue mold cheese	fried	any
t_{26}	venison	steamed	any	t_{78}	blue mold cheese	stewed	any
t_{27}	venison	boiled	any	t_{79}	blue mold cheese	casserole	any
t_{28}	venison	grilled	any	t_{80}	blue mold cheese	deep fried	any
t_{29}	venison	fried	any	t_{81}	fruit dessert	non (raw)	any
t_{30}	venison	stewed	any	t_{82}	fruit dessert	steamed	any
t_{31}	venison	casserole	any	t_{83}	fruit dessert	boiled	any
t_{32}	venison	deep fried	any	t_{84}	fruit dessert	grilled	any
t_{33}	fowl	non (raw)	any	t_{85}	fruit dessert	fried	any
t_{34}	fowl	steamed	any	t_{86}	fruit dessert	stewed	any
t_{35}	fowl	boiled	any	t_{87}	fruit dessert	casserole	any
t_{36}	fowl	grilled	any	t_{88}	fruit dessert	deep fried	any
t_{37}	fowl	fried	any	t_{89}	aromatic dessert	non (raw)	any
t_{38}	fowl	stewed	any	t_{90}	aromatic dessert	steamed	any
t_{39}	fowl	casserole	any	t_{91}	aromatic dessert	boiled	any

t_{40}	fowl	deep fried	any	t_{92}	aromatic dessert	grilled	any
t_{41}	fish	non (raw)	any	t_{93}	aromatic dessert	fried	any
t_{42}	fish	steamed	any	t_{94}	aromatic dessert	stewed	any
t_{43}	fish	boiled	any	t_{95}	aromatic dessert	casserole	any
t_{44}	fish	grilled	any	t_{96}	aromatic dessert	deep fried	any
t_{45}	fish	fried	any	t_{97}	ice cream	non (raw)	any
t_{46}	fish	stewed	any	t_{98}	ice cream	steamed	any
t_{47}	fish	casserole	any	t_{99}	ice cream	boiled	any
t_{48}	fish	deep fried	any	t_{100}	ice cream	grilled	any
t_{49}	hard cheese	non (raw)	any	t_{101}	ice cream	fried	any
t_{50}	hard cheese	steamed	any	t_{102}	ice cream	stewed	any
t_{51}	hard cheese	boiled	any	t_{103}	ice cream	casserole	any
t_{52}	hard cheese	grilled	any	t_{104}	ice cream	deep fried	any

$(S_1^{krit} \times \{any\} \times \{any\})$

t_{105}	pork	any	any	t_{112}	soft cheese	any	any
t_{106}	beef	any	any	t_{113}	goat cheese	any	any
t_{107}	veal	any	any	t_{114}	blue mold cheese	any	any
t_{108}	venison	any	any	t_{115}	fruit dessert	any	any
t_{109}	fowl	any	any	t_{116}	aromatic dessert	any	any
t_{110}	fish	any	any	t_{117}	ice cream	any	any
t_{111}	hard cheese	any	any				

$(\{any\} \times \{any\} \times S_3^{krit})$

t_{118}	any	any	Asian
t_{119}	any	any	Western

$(S_1^{krit} \times \{any\} \times S_3^{krit})$

t_{120}	pork	any	Asian	t_{133}	pork	any	Western
t_{121}	beef	any	Asian	t_{134}	beef	any	Western
t_{122}	veal	any	Asian	t_{135}	veal	any	Western
t_{123}	venison	any	Asian	t_{136}	venison	any	Western
t_{124}	fowl	any	Asian	t_{137}	fowl	any	Western
t_{125}	fish	any	Asian	t_{138}	fish	any	Western
t_{126}	hard cheese	any	Asian	t_{139}	hard cheese	any	Western
t_{127}	soft cheese	any	Asian	t_{140}	soft cheese	any	Western
t_{128}	goat cheese	any	Asian	t_{141}	goat cheese	any	Western
t_{129}	blue mold cheese	any	Asian	t_{142}	blue mold cheese	any	Western
t_{130}	fruit dessert	any	Asian	t_{143}	fruit dessert	any	Western
t_{131}	aromatic dessert	any	Asian	t_{144}	aromatic dessert	any	Western

6.2.1.5 Minimizing the set of all potential test data

Since there are no ordering relations with in the values of each sensor data, minimizing is no issue in this example domain.

6.2.2 Generation of a Reasonable Set of Test Cases (*ReST* ⊆ *QuEST*)

In (Knauf et al. 2002) and (Knauf 2000) a criteria-based reduction technology is suggested. Here, we tend to consider the following **criteria**:

1. If a potential test case is semantically more general and subsumed by another one (for instance [pork, any, any] is more general than [pork, any, Asian]), it is removed, i.e. only the more specific one "survives" the reduction procedure.
2. If a test case is a meal that doesn't exist at all (for instance "grilled ice cream"), it is removed.
3. Meals which exists in only one of the styles Asian or Western (raw fish, e.g.), are only considered in this style, not in the other one.
4. Meals, which don't have a system's solution, don't become an element of ***ReST***.
5. Desserts and Cheese are not distinguished in Asian and Western style.

After applying these criteria the following test data forms the reasonable set of test cases ***ReST***:

t_1	pork	boiled	Asian	t_{22}	fish	steamed	Western
t_2	pork	grilled	any	t_{23}	fish	boiled	Asian
t_3	pork	fried	any	t_{24}	fish	grilled	any
t_4	pork	stewed	any	t_{25}	fish	fried	any
t_5	beef	boiled	Asian	t_{26}	fish	stewed	Asian
t_6	beef	grilled	any	t_{27}	fish	deep fried	Asian
t_7	beef	fried	any	t_{28}	hard cheese	non (raw)	Western
t_8	beef	stewed	any	t_{29}	hard cheese	casserole	Western
t_9	veal	boiled	any	t_{30}	hard cheese	deep fried	Western
t_{10}	veal	grilled	any	t_{31}	soft cheese	non (raw)	Western
t_{11}	veal	fried	any	t_{32}	soft cheese	casserole	Western
t_{12}	veal	stewed	any	t_{33}	soft cheese	deep fried	Western
t_{13}	venison	boiled	any	t_{34}	goat cheese	non (raw)	Western
t_{14}	venison	grilled	any	t_{35}	goat cheese	casserole	Western
t_{15}	venison	fried	any	t_{36}	goat cheese	deep fried	Western
t_{16}	venison	stewed	any	t_{37}	blue mold cheese	non (raw)	Western
t_{17}	fowl	boiled	any	t_{38}	blue mold cheese	casserole	Western
t_{18}	fowl	grilled	any	t_{39}	blue mold cheese	deep fried	Western
t_{19}	fowl	fried	any	t_{40}	fruit dessert	non (raw)	any
t_{20}	fowl	stewed	any	t_{41}	aromatic dessert	non (raw)	any
t_{21}	fish	non (raw)	Asian	t_{42}	ice cream	non (raw)	any

6.3 Application Conditions

6.3.1 Available Experts, Associates Cases, Scheduled Sessions

Available Resources were

- human experts e_1, e_2, and e_3

  e_1 Mrs. Keiko TSURUTA, Inzai, Japan, tsuruta@sie.dendai.ac.jp

***e*₂** Mr. Lukas FREISTEDT, Weimar, Germany, l.freistedt@genion.de
***e*₃** Prof. Toshifumi TSUKIYAMA, Yokohama, Japan, tsukiyama@sie.dendai.ac.jp

- the reasonable set of test cases ***ReST*** with 42 test cases { t_1, ...,t_{42} }, and
- (by the end of the experimentation) also the server application ***TestMeToo*** to perform the TURING Test (cf. next session).

The **desired outcome** are answers to the following questions

1. *Does the **VKB** contribute to the validation sessions in an increasing degree with an increasing number of validation sessions?*
2. *Does the **VKB** contribute valid knowledge (best rated solutions) in an increasing degree with an increasing number of validation sessions?*
3. *Does the **VKB** skim the human expertise in an increasing degree with an increasing number of validation sessions?*
4. *Do the **VESA**s really model their human origin in an increasing degree with an increasing number of validation sessions?*

- Each of these four experts solves the **42 test cases** above in **4 sessions with 28 test cases each** (i.e. some test cases repetitively).
- This leads to up to **448 test case solutions** (up to 112 in each session), up to 336 solutions (84 per session) are provided by the human experts and 112 solutions (28 per session) provided by the system.
- This leads up to **504 test case ratings**.

Of course, very often all agents (human experts and the system) provide the same solution to a given test data. Furthermore, it is quite unlikely that a (human) expert changes his opinion to the same case from one session to the next. Thus, the number of solutions to be rated is much lower than this upper borders considered above.
The session plan is as follows:

session #	e_1	e_3	e_3	$VESA_1$	$VESA_2$	$VESA_3$	*ReST*
1	+	+	+	-	-	-	$ReST^1 = \{ t_1 , ..., t_{28} \}$
2	⊕	+	+	+	-	-	$ReST^2 = \{ t_{15} , ..., t_{42} \}$
3	+	⊕	+	-	+	-	$ReST^3 = \{ t_1 , .., t_{14} , t_{29} , .., t_{42} \}$
4	+	+	⊕	-	-	+	$ReST^4 = \{ t_i : t_i \bmod 3 \neq 0 \}$

\+ takes part in the sessions
\- takes **not** part in the sessions
⊕ takes part in solving and rating session only for being compared with its ***VESA***

The result of the ***i*** *-th* session are VKB^i, $VESA_1^i$, $VESA_2^i$, and $VESA_3^i$.

For a fair evaluation of the usefulness of ***VKB***, the intersection of test data in ***VKB*** and ***ReST*** (***EK*** = external knowledge) needs to be considered in each session:

- $EK_1 = \varnothing \cap ReST^1 = \varnothing$ $\quad | EK_1 | = 0$
- $EK_2 = \Pi_1(VKB^1) \cap ReST^2 = \{t_{15} , ..., t_{28}\}$ $\quad | EK_2 | = 14$
- $EK_3 = \Pi_1(VKB^2) \cap ReST^3 = ReST^3$ $\quad | EK_3 | = 28$
- $EK_4 = \Pi_1(VKB^3) \cap ReST^4 = ReST^4$ $\quad | EK_4 | = 28$

For the evaluation of the scheduled four sessions we determine after each session (session # ***i***)

- the number a_i of cases from ***VKB*** $^{i-1}$, which were subject of the rating session and relate it to $|EK_i|$: $A_i := a_i / |EK_i|$
- the number b_i of cases from ***VKB*** $^{i-1}$, which provided the optimal (best rated) solution and relate it to $|EK_i|$: $B_i := b_i / |EK_i|$
- the number c_i of cases from ***VKB*** $^{i-1}$, for which a new solution has been introduced into VKB and relate it to $|EK_i|$: $C_i := c_i / |EK_i|$
- the number d_i of solutions and ratings, which are identical responses of e_{i-1} and $VESA_{i-1}$ and relate it to the number of required solutions and ratings: $D_i := d_i /$ ***number of expert responses altogether***

Answers to the vacant questions can now expressed as follows:

1. *Does a **VKB** contribute to the validation sessions in an increasing degree with an increasing number of validation sessions:* $A_4 > A_3 > A_2$?
2. *Does the **VKB** contribute valid knowledge (best rated solutions) in an increasing degree with an increasing number of validation sessions:* $B_4 > B_3 > B_2$?
3. *Does the **VKB** skim the human expertise in an increasing degree with an increasing number of validation sessions:* $C_4 < C_3 < C_2$?
4. *Do the **VESA**s really model their human origin in an increasing degree with an increasing number of validation sessions:* $D_4 > D_3 > D_2$?

6.3.2 Available Tools

One of our students developed a tool called ***TestMeToo*** (abbr. of **Test** Case Experi**Me**ntation **Too**l) to demonstrate the usability of the ***VKB*** and ***VESA*** approaches (Kurbad 2003).

TestMeToo is a server application that can be accessed by clients by usual Web browsers (like ***Internet Explorer*** ©, ***Netscape*** ©, …). ***TestMeToo*** is available at

http://www.virtual-land.de/zope/testmetoo

As this address indicates, ***TestMeToo*** is implemented with an Open Source tool called ***Zope*** © [7], a high-performance object-oriented platform for building dynamic Web applications ***Zope*** integrates a Web server, an FTP server, and an Object Oriented data bank and enjoys a GNU Public License.
Zope itself is implemented with ***Python*** © [8], an Open Source Pearl- or PHP – like script language for Web applications.
Zope is a stable and highly maintained product. It combines benefits of both commercial and Open Source software. There is a special license of the ***Zope*** Corporation called ***ZPL*** (Zope Public License): Every product which is developed for

[7] See http://www.virtual-land.de/zope and http://www.zope.org/Credits for more information.
[8] See http://python.org/ for more information.

Zope or within the ***Zope*** framework, is half owned by the developer and half owned by the ***Zope*** Corporation. The developer has the right to sell it or to provide it for free, but the ***Zope*** Corporation has the same right. Since developers sell commercial applications based on the Open Source framework, they have a high interest in providing a stable base. Thus, ***Zope*** developments are usually moderated and every developer follows the given (or discussed) standards.

The decision for an Open Source solution revealed the following pros and cons:

+ The software is free of charge.
+ The source goes through “many hands” and is evaluated frequently.
+ There are user groups, mailing lists etc. for discussions, suggestions of updates and bug fixes.
+ Other products may be interfaced (or even integrated) seamlessly, since the interfaces of the Open Source solutions can be adopted to new needs.
+ Often, comments on specific issues can be found in the code itself. Therefore, no external documentation is necessary.
+ It is an unwritten law in many cases that existing standards should be strictly followed. ***Python***, for instance, exports and imports standardized XML without any problems.
– The (external) documentation often stays behind the development of the code or is of bad quality. This is the case with ***Zope***, for example.
– The maintenance and test might be costly and inefficient.
– There is no guarantee, that a specific function of the software will still be available (and/or working) in the next version.
– Many developers also represent many different views on what should be done and how it should be done. Thus, discussions about minor problems sometimes paralyze the development. This is especially the case, if conflicting standards for the same problem exist.

Since this report aims at an indication for the usefulness of the ***VKB*** and ***VESA*** concepts, it focuses the Test Case Experimentation step of the validation methodology as introduced in (Knauf 2000, Knauf et al. 2002) and sketched in section 2 of this report. Other steps, i.e. the test case generation, the evaluation, the validity estimation and the system refinement are performed manually following the concepts of (Knauf 2000, Knauf et al. 2002).

Since we limit the access to the enrolled experts for each application (i.e. for each TURING test Experiment), ***TestMeToo*** starts with a request for the provided password (see **Figure 3**).

Besides the solving and rating session there is a so called *Solution Identification Session*, which serves to clear up, whether a user a user, who provided a new solution (which has not been in the process so far) really means something different from the solutions so far or just has a different name for the same thing.

The introductory screens to each session type explain the user the purpose of the session and the way to perform it with ***TestMeToo*** (see Figure 4, Figure 5 and Figure 6)

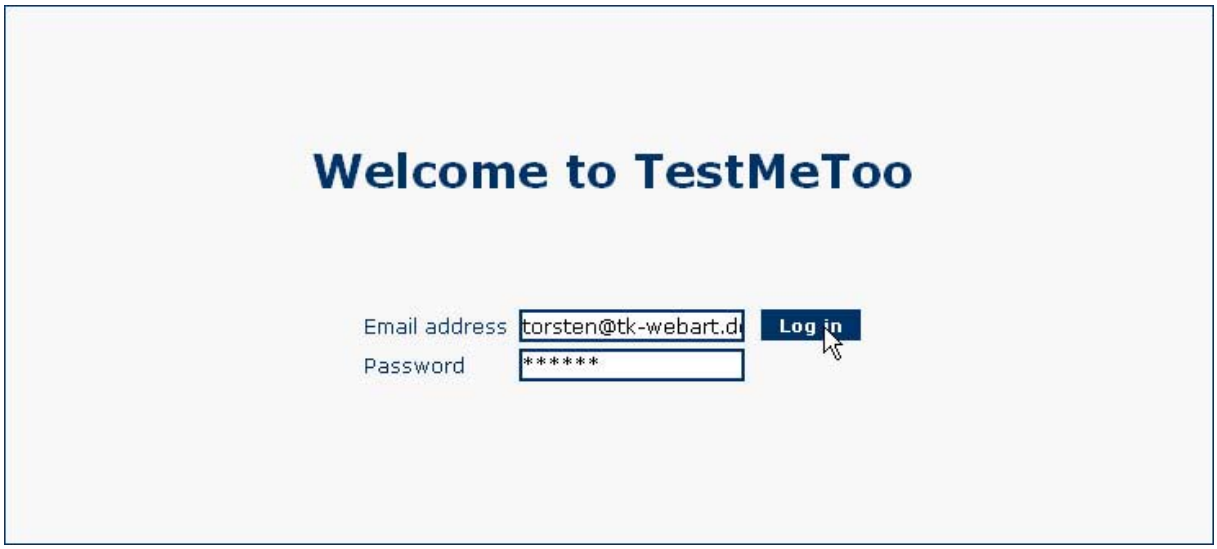

Figure 3: Login Screen to TestMeToo

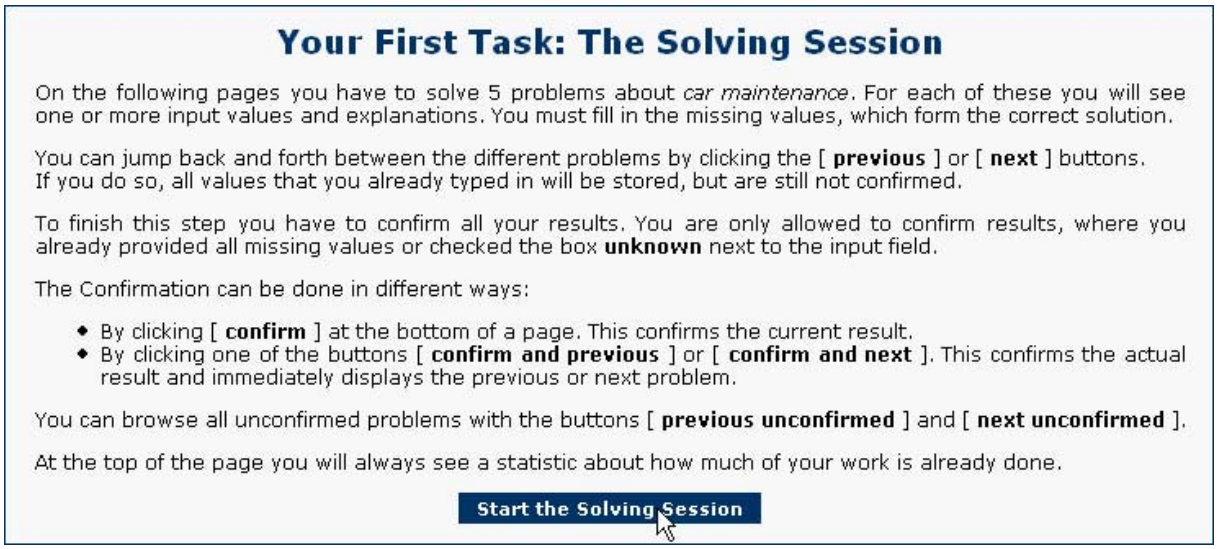

Figure 4: Explanation of the Solving Session

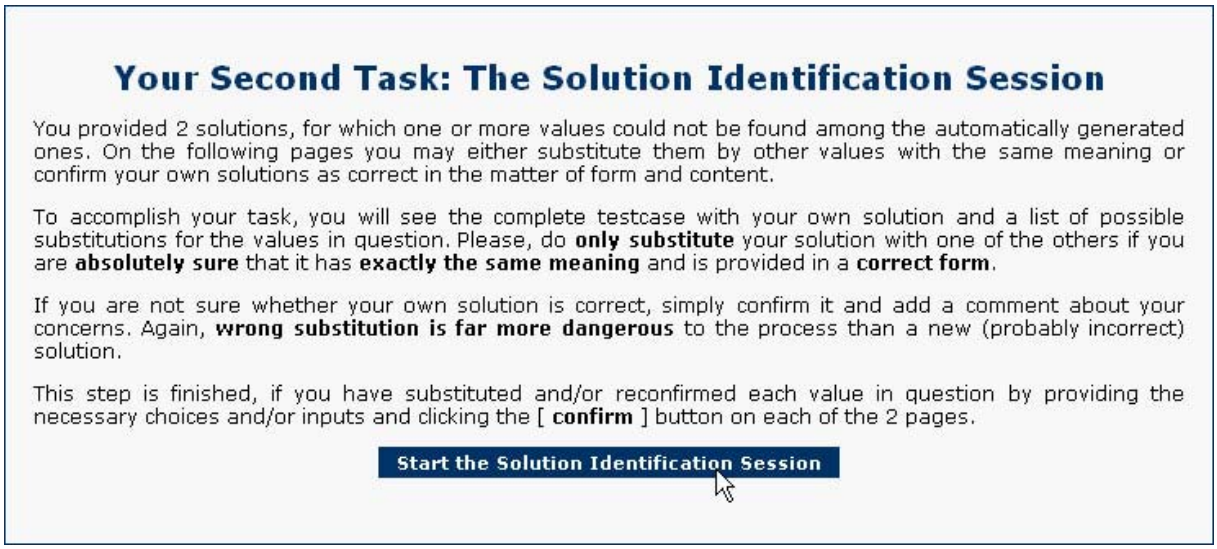

Figure 5: Explanation of the Solution Identification Session

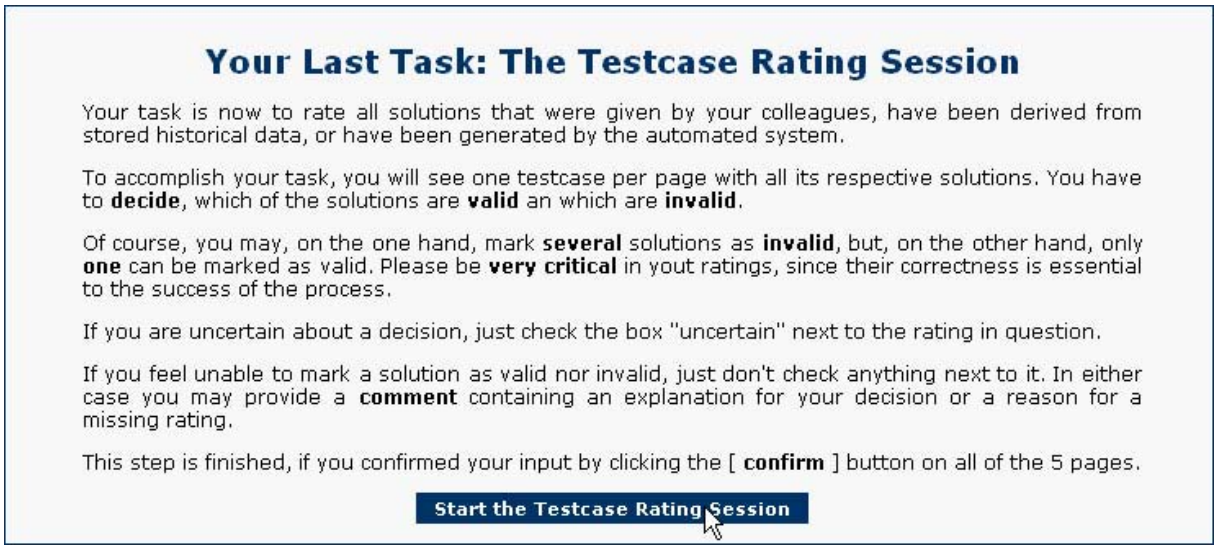

Figure 6: Explanation of the Rating Session

After the introduction to a solution session the user is asked for a solution that needs to be typed in the Solution Attribute slot (see Figure 7). Since this tool is also in use for an application with so called Business Rules, an output here consists (also) of a pair [Attribute,Value], just like the input attributes. The possible values for outputs in the experiment described here are just *yes* or *no*, so that we consider the attribute itself as the “real output”.
Furthermore, the kind of the particular inputs are explained (boolean, enumerable, integer, …) and the expert has the opportunity to drop a comment. This comment will be included in the informal description D_C of an entry in ***VKB***.

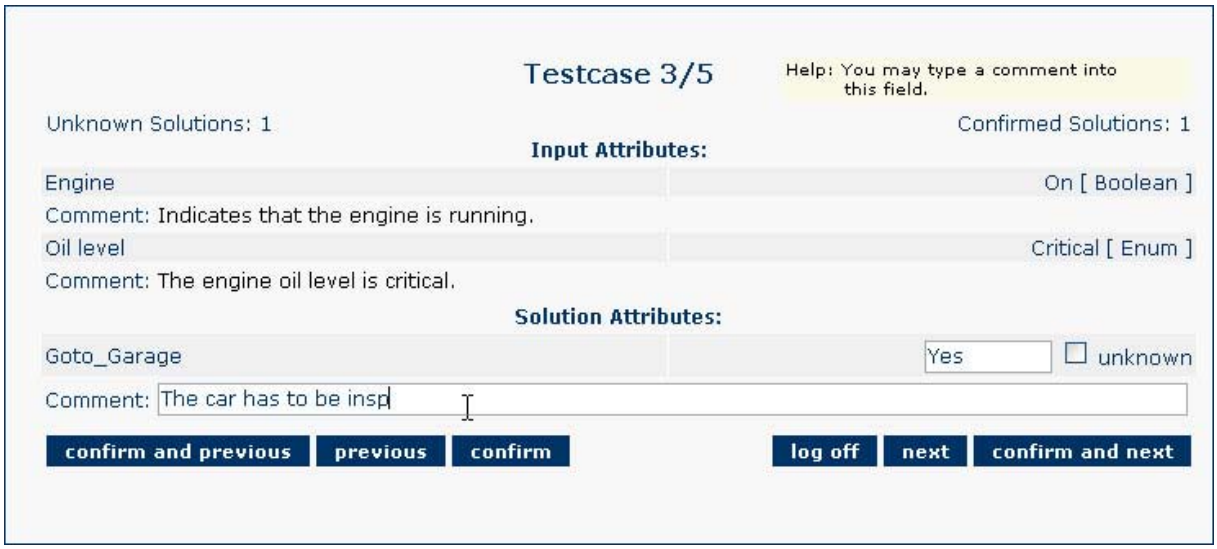

Figure 7: Solving Session

In the solution identification session the system lists up all provided solutions, which not identical to a system’s solution (see Figure 8). Here, the expert is asked, whether he/she really means the new solution he/she provided or a solution that is (semantically) identical to a one which is a system’s solution (for any test data). In fact, different syntactical characterizations for identical items is not solved satisfactory yet.

Figure 8: Solution Identification Session

An exemplarily screen shot for the rating session is shown in Figure 9. Again, with respect to the kind of rules in another application a solution is (also) an attribute, which needs a value.
Several solutions (attribute-value-parts, here) are listed up and the expert has the opportunity to click *valid* or *invalid* depending on his opinion about the correctness of the associated solution.

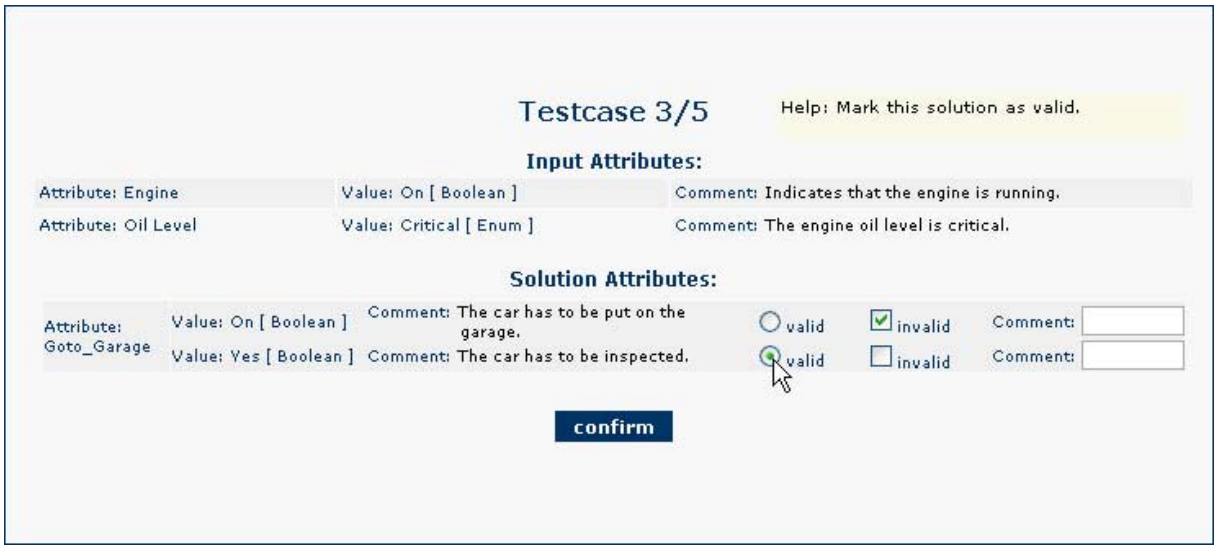

Figure 9: Rating Session

6.3.3 Running the Sessions

Since ***TestMeToo*** (cf. previous session) was under construction at the time the first session needed to start, a questionnaire for the first sessions (see Appendix A) were developed and mailed to the experts.

The list of system outputs is also presented as “possible solutions”, because the concept of identity of syntactically different verbal descriptions is not solved satisfactory yet. In fact, this problem might rise again with the new solutions that are provided by the experts.

The problem of non-determinism of the system behavior is considered to be solved by providing the solution of the first rule trace in the knowledge base, just as PROLOG does. as the system's solution.

Non-determinism in an expert's behavior (by providing multiple ratings) is tolerated, because a certain expert can support several solutions.

We omitted the system refinement step, because
1. the refinement is not essential to show the expected result with respect to the usefulness of VKB and VESA and
2. there is no software yet, that performs the refinement, i.e. the technique needs to be applied manually, which is beyond the time limit of the present experiment.

The test data itself raised some questions to the experts with respect to its interpretation. After a topical discussion we agreed to the following interpretation:
- ***Fish*** means "white fish", not red one.
- ***Cheese*** means the one with the strongest taste in its category, because cheese with a light taste would not be a "main ingredient".
- ***Asian style*** means the more spicy variant of it. This is the more Chinese style, less the Japanese one.
- ***Western style*** means no spice, a little salty. Meat in this style goes along with brown sauce (not white one) and in case of fish it is supplemented by garlic.
- ***Any style*** means no spice besides salt and pepper. In case stewed food it also means a dark colored sauce.

7 Test Results

7.1 1st session

As a result
- of filling in the questionnaire shown in 10.1 by the three experts e_1, e_2, and e_3 and
- running the 28 test cases by the system

we received the following outputs o_k provided by the experts e_i results to the test data t_j. New solutions (different from all system's solution) did not come up.

	e_1	e_2	e_3	system		e_1	e_2	e_3	system
t_1	o_6	o_{18}	o_6	o_{18}	t_{15}	o_{10}	o_9	o_3	o_3
t_2	o_7	o_{21}	o_7	o_2	t_{16}	o_{11}	o_9	o_{11}	o_{10}
t_3	o_8	o_{20}	o_7	o_2	t_{17}	o_1	o_{18}	o_{20}	o_1
t_4	o_2	o_1	o_2	o_{10}	t_{18}	o_6	o_{20}	o_6	o_1
t_5	o_8	o_3	o_8	o_{18}	t_{19}	o_{12}	o_{21}	o_{12}	o_1
t_6	o_9	o_9	o_9	o_2	t_{20}	o_1	o_{21}	o_{20}	o_1
t_7	o_3	o_3	o_3	o_2	t_{21}	o_{16}	o_{21}	o_{14}	o_{18}
t_8	o_4	o_3	o_3	o_{10}	t_{22}	o_{17}	o_{14}	o_{17}	o_{14}
t_9	o_1	o_{20}	o_{13}	o_1	t_{23}	o_{13}	o_{16}	o_{15}	o_{18}
t_{10}	o_2	o_{12}	o_{12}	o_1	t_{24}	o_{19}	o_{16}	o_{19}	o_{15}
t_{11}	o_{13}	o_4	o_{12}	o_1	t_{25}	o_{20}	o_{16}	o_{20}	o_{14}
t_{12}	o_5	o_4	o_1	o_1	t_{26}	o_{16}	o_{16}	o_{14}	o_{18}
t_{13}	o_8	o_9	o_9	o_3	t_{27}	o_{18}	o_{16}	o_{13}	o_{18}

	e_1	e_2	e_3	system		e_1	e_2	e_3	system
t_{14}	o_9	o_9	o_3	o_3	t_{28}	o_9	o_8	o_{16}	o_{16}

This led to the questionnaire for the first rating session as listed in section 10.2. The rating session based on this questionnaire led to the following result. The own solution of each expert is marked red. In the test cases marked with green an additional solution without any (human or machine) source was added in the questionnaire by mistake; fortunately, this does not influence the result.

test data	solution	e_1		e_2		e_3	
		rating	certainty	rating	certainty	rating	certainty
t_1	o_6	1	0	0	1	1	1
t_1	o_{18}	0	0	1	1	0	1
t_1	o_{19}	1	0	0	0	0	1
t_2	o_2	1	0	0	1	1	0
t_2	o_7	0	0	0	0	1	1
t_2	o_{15}	0	0	1	1	0	1
t_2	o_{21}	0	0	1	1	0	1
t_3	o_2	1	0	0	1	1	1
t_3	o_8	1	0	0	1	1	0
t_3	o_7	0	0	1	1	0	1
t_3	o_{20}	0	0	1	1	0	1
t_4	o_1	1	0	1	1	1	0
t_4	o_2	0	0	1	1	1	1
t_4	o_{10}	1	0	0	0	1	0
t_4	o_{20}	0	0	1	1	0	1
t_5	o_1	1	0	1	0	1	1
t_5	o_3	1	1	1	0	0	1
t_5	o_8	1	1	1	0	1	1
t_5	o_{18}	0	0	0	0	0	1
t_6	o_2	0	0	1	0	1	1
t_6	o_3	0	0	1	1	1	1
t_6	o_9	1	0	1	1	1	1
t_7	o_2	0	0	1	1	1	0
t_7	o_3	1	0	1	1	1	1
t_8	o_3	1	0	1	1	1	1
t_8	o_4	1	0	1	1	0	1
t_8	o_5	1	0	1	1	0	1
t_8	o_{10}	1	0	0	0	0	0
t_9	o_1	1	0	1	1	1	1
t_9	o_{13}	0	0	0	1	0	1
t_9	o_{20}	0	0	0	1	0	1
t_{10}	o_1	1	0	1	1	1	1
t_{10}	o_2	1	0	1	1	0	1
t_{10}	o_{12}	1	0	1	1	1	1
t_{11}	o_1	1	1	1	1	1	1
t_{11}	o_4	1	0	1	1	1	1

test data	solution	e_1		e_2		e_3	
		rating	certainty	rating	certainty	rating	certainty
t_{11}	o_{12}	1	0	1	1	**1**	**1**
t_{11}	o_{13}	**0**	**0**	0	1	0	1
t_{12}	o_1	1	1	1	1	**1**	**1**
t_{12}	o_4	1	0	**1**	**0**	1	1
t_{12}	o_5	**1**	**0**	1	1	1	1
t_{13}	o_2	0	0	1	1	0	0
t_{13}	o_3	1	0	1	1	1	0
t_{13}	o_8	**1**	**0**	1	1	1	1
t_{13}	o_9	1	0	**1**	**1**	**1**	**1**
t_{14}	o_3	0	0	1	1	**1**	**1**
t_{14}	o_9	**1**	**0**	**1**	**1**	1	1
t_{15}	o_3	0	0	1	1	**1**	**1**
t_{15}	o_9	1	0	**1**	**1**	1	0
t_{15}	o_{10}	**1**	**1**	0	0	1	0
t_{16}	o_2	0	0	1	1	1	0
t_{16}	o_9	1	1	**1**	**1**	1	1
t_{16}	o_{10}	1	0	0	0	1	1
t_{16}	o_{11}	**1**	**0**	0	0	**1**	**1**
t_{17}	o_1	**0**	**0**	0	1	1	1
t_{17}	o_{18}	0	0	**1**	**0**	0	1
t_{17}	o_{20}	1	0	1	1	**1**	**1**
t_{18}	o_1	0	0	0	1	1	1
t_{18}	o_6	**0**	**0**	1	1	**1**	**1**
t_{18}	o_{13}	1	0	0	0	0	1
t_{18}	o_{20}	1	0	**1**	**1**	0	1
t_{19}	o_1	0	0	0	1	1	1
t_{19}	o_{12}	**1**	**0**	1	1	**1**	**1**
t_{19}	o_{13}	1	0	0	0	0	1
t_{19}	o_{21}	1	0	**1**	**1**	0	1
t_{20}	o_1	**0**	**0**	0	0	1	1
t_{20}	o_{20}	1	0	1	1	**0**	**1**
t_{20}	o_{21}	1	0	**1**	**1**	0	1
t_{21}	o_{14}	1	0	1	1	**0**	**1**
t_{21}	o_{16}	**0**	**0**	1	1	1	1
t_{21}	o_{18}	0	0	1	1	1	1
t_{21}	o_{21}	1	1	**1**	**1**	1	1
t_{22}	o_{14}	1	0	**1**	**1**	1	0
t_{22}	o_{17}	**1**	**1**	1	1	**1**	**1**
t_{23}	o_{13}	**1**	**0**	0	0	0	1
t_{23}	o_{15}	1	0	1	0	**1**	**1**
t_{23}	o_{16}	1	0	**1**	**0**	1	1
t_{23}	o_{18}	0	0	1	1	0	1
t_{24}	o_{15}	1	0	1	1	1	1
t_{24}	o_{16}	1	1	**1**	**0**	1	1
t_{24}	o_{17}	1	0	1	1	0	1

test data	solution	e_1		e_2		e_3	
		rating	certainty	rating	certainty	rating	certainty
t_{24}	o_{19}	0	0	0	0	1	1
t_{25}	o_{13}	0	0	0	0	0	1
t_{25}	o_{14}	1	0	1	1	1	1
t_{25}	o_{16}	1	0	1	1	1	1
t_{25}	o_{20}	1	1	1	1	1	1
t_{26}	o_{14}	1	0	0	0	0	1
t_{26}	o_{16}	1	0	1	1	1	1
t_{26}	o_{18}	0	0	1	1	0	1
t_{27}	o_{13}	0	0	0	0	0	0
t_{27}	o_{16}	1	0	1	1	1	1
t_{27}	o_{18}	0	0	1	0	1	1
t_{28}	o_3	1	0	1	1	1	1
t_{28}	o_8	1	0	1	1	1	1
t_{28}	o_9	1	0	1	1	1	1
t_{28}	o_{16}	0	0	0	1	1	1

The competences of each expert per test case following the approach of (Knauf 2000) and the optimal solution that gained the best certain rating (weighted by the experts' competences each) of the expert panel are as follows (represented as fractions for exactness). In case concurrent solutions enjoyed optimality, the sum of the competences of the (also uncertain) supporters decides. If this is the same, the sum of the competences of solvers decides.
In case this is still the same number, we used here the overall competence (average over all test cases). The solution provided by the most competent expert over all test cases enjoys optimality.
If this is still not a KO criterion, the systems solution is preferred.

test data	$cpt\ (e_1, t_j)$	$cpt\ (e_2, t_j)$	$cpt\ (e_3, t_j)$	sol_{Kj}^{opt}
t_1	1 / 2	7 / 12	2 / 3	o_6
t_2	1 / 2	11 / 18	11 / 18	o_7
t_3	1 / 3	2 / 3	11 / 18	o_{20}
t_4	1 / 2	11 / 18	8 / 9	o_2
t_5	8 / 9	1 / 2	1	o_8
t_6	2 / 3	11 / 12	1	o_9
t_7	2 / 3	1	5 / 6	o_3
t_8	1 / 2	17 / 18	17 / 18	o_3
t_9	2 / 3	1 / 2	1 / 2	o_1
t_{10}	1 / 2	1	1	o_1
t_{11}	1 / 6	1	1	o_1
t_{12}	3 / 4	5 / 6	1	o_1
t_{13}	2 / 3	1	8 / 9	o_9
t_{14}	2 / 3	1	1	o_3
t_{15}	1 / 2	7 / 12	5 / 6	o_3
t_{16}	13 / 18	8 / 9	11 / 18	o_9
t_{17}	1 / 3	1 / 2	1	o_{20}
t_{18}	1 / 2	11 / 18	1	o_6

test data	$cpt\ (e_1, t_j)$	$cpt\ (e_2, t_j)$	$cpt\ (e_3, t_j)$	sol_{Kj}^{opt}
t_{19}	2 / 3	11 / 18	1	o_{12}
t_{20}	1 / 2	7 / 12	5 / 6	o_{21}
t_{21}	13 / 18	2 / 3	5 / 6	o_{21}
t_{22}	5 / 6	2 / 3	5 / 6	o_{17}
t_{23}	1 / 3	8 / 9	2 / 3	o_{15}
t_{24}	1 / 2	7 / 9	2 / 3	o_{15}
t_{25}	5 / 6	17 / 18	1	o_{20}
t_{26}	2 / 3	11 / 12	1 / 2	o_{16}
t_{27}	1 / 2	5 / 6	1 / 3	o_{16}
t_{28}	2 / 3	1	2 / 3	o_8
overall competence	16,25	21,64	22,72	

The resulting VKB^1 is

t_j	E_K	E_I	sol_{Kj}^{opt}	r_{ijk}	c_{ijk}	τ_s	D_C
t_1	e_1, e_3	$[e_1, e_2, e_3]$	o_6	[1,0,1]	[0,1,1]	1	
t_2	e_1, e_3	$[e_1, e_2, e_3]$	o_7	[0,0,1]	[0,0,1]	1	
t_3	e_2	$[e_1, e_2, e_3]$	o_{20}	[0,1,0]	[0,1,1]	1	
t_4	e_1, e_3	$[e_1, e_2, e_3]$	o_2	[0,1,1]	[0,1,1]	1	
t_5	e_1, e_3	$[e_1, e_2, e_3]$	o_8	[1,1,1]	[1,0,1]	1	
t_6	e_1, e_2, e_3	$[e_1, e_2, e_3]$	o_9	[1,1,1]	[0,1,1]	1	
t_7	e_1, e_2, e_3	$[e_1, e_2, e_3]$	o_3	[1,1,1]	[0,1,1]	1	
t_8	e_2, e_3	$[e_1, e_2, e_3]$	o_3	[1,1,1]	[0,1,1]	1	
t_9	e_1	$[e_1, e_2, e_3]$	o_1	[1,1,1]	[0,1,1]	1	
t_{10}	∅	$[e_1, e_2, e_3]$	o_1	[1,1,1]	[0,1,1]	1	
t_{11}	∅	$[e_1, e_2, e_3]$	o_1	[1,1,1]	[1,1,1]	1	
t_{12}	e_3	$[e_1, e_2, e_3]$	o_1	[1,1,1]	[1,1,1]	1	
t_{13}	e_2, e_3	$[e_1, e_2, e_3]$	o_9	[1,1,1]	[0,1,1]	1	
t_{14}	e_3	$[e_1, e_2, e_3]$	o_3	[0,1,1]	[0,1,1]	1	
t_{15}	e_3	$[e_1, e_2, e_3]$	o_3	[0,1,1]	[0,1,1]	1	
t_{16}	e_2	$[e_1, e_2, e_3]$	o_9	[1,1,1]	[1,1,1]	1	
t_{17}	e_3	$[e_1, e_2, e_3]$	o_{20}	[1,1,1]	[0,1,1]	1	
t_{18}	e_1, e_3	$[e_1, e_2, e_3]$	o_6	[0,1,1]	[0,1,1]	1	
t_{19}	e_1, e_3	$[e_1, e_2, e_3]$	o_{12}	[1,1,1]	[0,1,1]	1	
t_{20}	e_2	$[e_1, e_2, e_3]$	o_{21}	[1,1,0]	[0,1,1]	1	
t_{21}	e_2	$[e_1, e_2, e_3]$	o_{21}	[1,1,1]	[1,1,1]	1	
t_{22}	e_1, e_3	$[e_1, e_2, e_3]$	o_{17}	[1,1,1]	[1,1,1]	1	
t_{23}	e_3	$[e_1, e_2, e_3]$	o_{15}	[1,1,1]	[0,0,1]	1	
t_{24}	∅	$[e_1, e_2, e_3]$	o_{15}	[1,1,1]	[0,1,1]	1	
t_{25}	e_1, e_3	$[e_1, e_2, e_3]$	o_{20}	[1,1,1]	[1,1,1]	1	
t_{26}	e_1, e_2	$[e_1, e_2, e_3]$	o_{16}	[1,1,1]	[0,1,1]	1	
t_{27}	e_2	$[e_1, e_2, e_3]$	o_{16}	[1,1,1]	[0,1,1]	1	
t_{28}	e_2	$[e_1, e_2, e_3]$	o_8	[1,1,1]	[0,1,1]	1	
t_{29} … t_{42}	∅	∅	no solution	no rating	0	1	∅

7.2 2nd session

As a result
- of filling in the questionnaire shown in 10.3 by the three experts e_1, e_2, and e_3 and
- running the 28 test cases by the system

we received the following outputs o_k provided by the experts e_i results to the test data t_j. New solutions (different from all system's solutions) did not come up.

	e_1	e_2	e_3	system		e_1	e_2	e_3	system
t_{15}	o_3	o_9	o_3	o_3	t_{29}	o_{15}	o_8	o_{16}	o_{16}
t_{16}	o_3	o_9	o_{11}	o_{11}	t_{30}	o_{16}	o_9	o_{16}	o_{16}
t_{17}	o_{14}	o_{18}	o_{20}	o_1	t_{31}	o_2	o_2	o_2	o_{16}
t_{18}	o_{18}	o_{20}	o_6	o_1	t_{32}	o_4	o_3	o_2	o_{16}
t_{19}	o_{13}	o_{21}	o_{12}	o_1	t_{33}	o_3	o_8	o_2	o_{16}
t_{20}	o_{12}	o_{21}	o_4	o_1	t_{34}	o_9	o_2	o_5	o_{17}
t_{21}	o_{15}	o_{21}	o_{22}	o_{18}	t_{35}	o_{10}	o_8	o_5	o_{17}
t_{22}	o_{17}	o_{14}	o_{18}	o_{14}	t_{36}	o_{11}	o_9	o_5	o_{17}
t_{23}	o_{13}	o_{16}	o_{20}	o_{18}	t_{37}	o_2	o_9	o_{10}	o_{24}
t_{24}	o_{19}	o_{16}	o_{19}	o_{15}	t_{38}	o_6	o_9	o_{11}	o_{24}
t_{25}	o_{20}	o_{16}	o_{20}	o_{15}	t_{39}	o_9	o_9	o_9	o_{24}
t_{26}	o_{21}	o_{16}	o_{20}	o_{18}	t_{40}	o_{23}	o_{23}	o_{23}	o_{22}
t_{27}	o_{22}	o_{16}	o_{22}	o_{18}	t_{41}	o_{19}	o_{22}	o_{18}	o_{19}
t_{28}	o_6	o_8	o_{16}	o_{16}	t_{42}	o_{23}	o_{23}	o_{23}	o_{23}

This led to the questionnaire for the first rating session as listed in section 10.4. The rating session based on this questionnaire led to the following result. The own solution of each expert is marked red. The test data marked blue has been introduced from ***VKB***[1].

test data	solution	e_1		e_2		e_3	
		rating	certainty	rating	certainty	rating	certainty
t_{15}	o_3	**1**	**1**	1	1	**1**	**1**
t_{15}	o_9	1	1	**1**	**1**	1	1
t_{16}	o_3	**1**	**1**	1	1	1	1
t_{16}	o_9	1	1	**1**	**1**	1	1
t_{16}	o_{11}	1	1	1	0	**1**	**1**
t_{17}	o_1	0	1	0	1	1	1
t_{17}	o_{14}	**1**	**1**	1	1	0	1
t_{17}	o_{18}	0	1	**0**	**0**	0	1
t_{17}	o_{20}	1	1	1	1	**1**	**0**
t_{18}	o_1	0	1	0	1	1	1
t_{18}	o_6	0	1	1	1	**1**	**1**
t_{18}	o_{18}	**0**	**1**	1	0	0	1
t_{18}	o_{20}	1	1	**1**	**1**	0	0
t_{19}	o_1	0	1	0	1	1	1
t_{19}	o_{12}	1	1	0	1	**1**	**1**
t_{19}	o_{13}	**1**	**1**	1	0	0	1
t_{19}	o_{21}	0	1	**1**	**1**	0	1

test data	solution	e_1		e_2		e_3	
		rating	certainty	rating	certainty	rating	certainty
t_{20}	o_1	0	1	0	1	1	0
t_{20}	o_4	1	1	0	1	**1**	**1**
t_{20}	o_{12}	**1**	**1**	0	1	1	1
t_{20}	o_{21}	1	1	**1**	**1**	0	1
t_{21}	o_{15}	**1**	**1**	1	1	1	1
t_{21}	o_{18}	0	1	1	0	1	0
t_{21}	o_{21}	1	1	**1**	**1**	1	0
t_{21}	o_{22}	1	1	1	1	**1**	**1**
t_{22}	o_{14}	1	1	**1**	**1**	1	1
t_{22}	o_{17}	**1**	**1**	1	1	1	1
t_{22}	o_{18}	0	1	0	0	**1**	**0**
t_{23}	o_{13}	**1**	**1**	1	0	1	0
t_{23}	o_{15}	1	0	1	0	1	0
t_{23}	o_{16}	1	1	**1**	**1**	1	1
t_{23}	o_{18}	0	1	0	0	1	1
t_{23}	o_{20}	1	1	1	1	**1**	**1**
t_{24}	o_{15}	1	1	1	1	1	1
t_{24}	o_{16}	1	1	**1**	**1**	1	1
t_{24}	o_{19}	**0**	**1**	0	0	**1**	**1**
t_{25}	o_{15}	1	1	1	1	1	1
t_{25}	o_{16}	1	1	**1**	**1**	1	1
t_{25}	o_{20}	**1**	**1**	1	1	**1**	**1**
t_{26}	o_{16}	1	1	**1**	**1**	1	1
t_{26}	o_{18}	1	1	0	0	1	0
t_{26}	o_{20}	1	1	1	1	**1**	**1**
t_{26}	o_{21}	**1**	**1**	1	1	1	1
t_{27}	o_{16}	1	1	**1**	**1**	1	1
t_{27}	o_{18}	0	1	0	0	1	1
t_{27}	o_{22}	**1**	**1**	1	1	**1**	**1**
t_{28}	o_6	**1**	**1**	1	0	1	1
t_{28}	o_8	1	1	**1**	**1**	1	1
t_{28}	o_{16}	0	1	0	1	**1**	**0**
t_{29}	o_8	1	1	**1**	**1**	1	0
t_{29}	o_{15}	**0**	**1**	0	1	0	0
t_{29}	o_{16}	0	1	0	1	**0**	**0**
t_{30}	o_9	1	1	**1**	**1**	1	1
t_{30}	o_{16}	**0**	**1**	0	1	**0**	**1**
t_{31}	o_2	**0**	**1**	**1**	**1**	**1**	**1**
t_{31}	o_{16}	0	1	0	1	1	1
t_{32}	o_2	1	1	1	1	**1**	**0**
t_{32}	o_3	1	1	**1**	**1**	1	1
t_{32}	o_4	**1**	**1**	1	1	1	1
t_{32}	o_{16}	1	1	0	1	1	0
t_{33}	o_2	1	1	1	1	**1**	**1**
t_{33}	o_3	**1**	**1**	1	1	1	1

test data	solution	e_1		e_2		e_3	
		rating	certainty	rating	certainty	rating	certainty
t_{33}	o_8	1	1	**1**	**1**	1	1
t_{33}	o_{16}	1	1	0	1	0	0
t_{34}	o_2	1	1	**1**	**1**	1	1
t_{34}	o_5	0	1	0	1	**1**	**0**
t_{34}	o_9	**1**	**1**	1	1	1	1
t_{34}	o_{17}	0	1	0	1	0	1
t_{35}	o_5	0	1	0	1	**1**	**1**
t_{35}	o_8	1	1	**1**	**1**	1	1
t_{35}	o_{10}	**1**	**1**	1	1	1	1
t_{35}	o_{17}	0	1	0	1	0	1
t_{36}	o_5	0	1	0	1	**1**	**1**
t_{36}	o_9	1	1	**1**	**1**	1	1
t_{36}	o_{11}	**1**	**1**	1	1	1	1
t_{36}	o_{17}	0	1	0	1	0	0
t_{37}	o_2	**1**	**1**	1	1	1	1
t_{37}	o_9	1	1	**1**	**1**	1	1
t_{37}	o_{10}	1	1	1	1	**1**	**1**
t_{37}	o_{24}	0	1	0	1	0	0
t_{38}	o_6	**1**	**1**	1	1	1	1
t_{38}	o_9	1	1	**1**	**1**	1	1
t_{38}	o_{11}	0	1	1	1	**1**	**1**
t_{38}	o_{24}	0	1	0	1	0	0
t_{39}	o_9	**1**	**1**	**1**	**1**	**1**	**1**
t_{39}	o_{24}	0	1	0	1	0	0
t_{40}	o_{22}	0	1	0	1	1	1
t_{40}	o_{23}	**1**	**1**	**1**	**1**	**1**	**1**
t_{41}	o_{18}	1	1	1	1	**1**	**1**
t_{41}	o_{19}	**1**	**1**	1	1	1	1
t_{41}	o_{22}	0	1	**0**	**1**	0	1
t_{42}	o_{23}	**1**	**1**	**1**	**1**	**1**	**1**

The competences of each expert per test case following the approach of (Knauf 2000) and the optimal solution that gained the best certain rating (weighted by the experts' competences each) of the expert panel are as follows (represented as fractions for exactness). In case concurrent solutions enjoyed optimality, the sum of the competences of the (also uncertain) supporters decides. If this is the same, the sum of the competences of solvers decides.
In case this is still the same number, we used here the overall competence (average over all test cases). The solution provided by the most competent expert over all test cases enjoys optimality.
If this is still not a KO criterion, the systems solution is preferred.

test data	$cpt\ (e_1, t_j)$	$cpt\ (e_2, t_j)$	$cpt\ (e_3, t_j)$	sol_{Kj}^{opt}
t_{15}	1	1	1	o_3
t_{16}	1	11 / 12	1	o_3
t_{17}	5 / 6	2 / 3	5 / 6	o_{20}

test data	$cpt\ (e_1, t_j)$	$cpt\ (e_2, t_j)$	$cpt\ (e_3, t_j)$	sol_{Kj}^{opt}
t_{18}	1 / 2	17 / 18	7 / 9	o_6
t_{19}	2 / 3	11 / 18	5 / 6	o_{12}
t_{20}	5 / 6	5 / 6	7 / 9	o_4
t_{21}	1	17 / 18	17 / 18	o_{15}
t_{22}	1	11 / 12	1 / 2	o_{17}
t_{23}	2 / 3	8 / 9	17 / 18	o_{20}
t_{24}	5 / 6	11 / 12	2 / 3	o_{16}
t_{25}	1	1	1	o_{20}
t_{26}	1	17 / 18	17 / 18	o_{21}
t_{27}	1	11 / 12	1	o_{22}
t_{28}	1	11 / 12	1 / 2	o_8
t_{29}	1 / 2	1	1 / 6	o_8
t_{30}	1 / 2	1	1 / 2	o_9
t_{31}	5 / 6	5 / 6	5 / 6	o_2
t_{32}	1	1	7 / 9	o_3
t_{33}	1	1	17 / 18	o_8
t_{34}	1	1	1 / 2	o_2
t_{35}	1	1	2 / 3	o_8
t_{36}	1	1	11 / 18	o_9
t_{37}	1	1	17 / 18	o_9
t_{38}	1	1	7 / 9	o_9
t_{39}	1	1	5 / 6	o_9
t_{40}	1	1	1	o_{23}
t_{41}	1	1 / 2	1	o_{19}
t_{42}	1	1	1	o_{23}
overall competence	25,17	25,14	22,28	

The resulting ***VKB²*** is

t_j	E_K	E_I	sol_{Kj}^{opt}	r_{ijk}	c_{ijk}	τ_s	D_C
t_1	e_1, e_3	$[e_1, e_2, e_3]$	o_6	[1,0,1]	[0,1,1]	1	
t_2	e_1, e_3	$[e_1, e_2, e_3]$	o_7	[0,0,1]	[0,0,1]	1	
t_3	e_2	$[e_1, e_2, e_3]$	o_{20}	[0,1,0]	[0,1,1]	1	
t_4	e_1, e_3	$[e_1, e_2, e_3]$	o_2	[0,1,1]	[0,1,1]	1	
t_5	e_1, e_3	$[e_1, e_2, e_3]$	o_8	[1,1,1]	[1,0,1]	1	
t_6	e_1, e_2, e_3	$[e_1, e_2, e_3]$	o_9	[1,1,1]	[0,1,1]	1	
t_7	e_1, e_2, e_3	$[e_1, e_2, e_3]$	o_3	[1,1,1]	[0,1,1]	1	
t_8	e_2, e_3	$[e_1, e_2, e_3]$	o_3	[1,1,1]	[0,1,1]	1	
t_9	e_1	$[e_1, e_2, e_3]$	o_1	[1,1,1]	[0,1,1]	1	
t_{10}	$\varnothing$	$[e_1, e_2, e_3]$	o_1	[1,1,1]	[0,1,1]	1	
t_{11}	$\varnothing$	$[e_1, e_2, e_3]$	o_1	[1,1,1]	[1,1,1]	1	
t_{12}	e_3	$[e_1, e_2, e_3]$	o_1	[1,1,1]	[1,1,1]	1	
t_{13}	e_2, e_3	$[e_1, e_2, e_3]$	o_9	[1,1,1]	[0,1,1]	1	
t_{14}	e_3	$[e_1, e_2, e_3]$	o_3	[0,1,1]	[0,1,1]	1	
t_{15}	e_3	$[e_1, e_2, e_3]$	o_3	[0,1,1]	[0,1,1]	1	
	e_1, e_3	$[e_1, e_2, e_3]$	o_3	[1,1,1]	[1,1,1]	2	

t_j	E_K	E_I	sol_{Kj}^{opt}	r_{ijk}	c_{ijk}	τ_s	D_C
t_{16}	e_2	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,1]	1	
	e_1	[e_1 , e_2 , e_3]	o_3	[1,1,1]	[1,1,1]	2	
t_{17}	e_3	[e_1 , e_2 , e_3]	o_{20}	[1,1,1]	[0,1,1]	1	
	e_3	[e_1 , e_2 , e_3]	o_{20}	[1,1,1]	[1,1,0]	2	
t_{18}	e_1 , e_3	[e_1 , e_2 , e_3]	o_6	[0,1,1]	[0,1,1]	1	
	e_3	[e_1 , e_2 , e_3]	o_6	[0,1,1]	[1,1,1]	2	
t_{19}	e_1 , e_3	[e_1 , e_2 , e_3]	o_{12}	[1,1,1]	[0,1,1]	1	
	e_3	[e_1 , e_2 , e_3]	o_{12}	[1,0,1]	[1,1,1]	2	
t_{20}	e_2	[e_1 , e_2 , e_3]	o_{21}	[1,1,0]	[0,1,1]	1	
	e_3	[e_1 , e_2 , e_3]	o_4	[1,0,1]	[1,1,1]	2	
t_{21}	e_2	[e_1 , e_2 , e_3]	o_{21}	[1,1,1]	[1,1,1]	1	
	e_1	[e_1 , e_2 , e_3]	o_{15}	[1,1,1]	[1,1,1]	2	
t_{22}	e_1 , e_3	[e_1 , e_2 , e_3]	o_{17}	[1,1,1]	[1,1,1]	1	
	e_1	[e_1 , e_2 , e_3]	o_{17}	[1,1,1]	[1,1,1]	2	
t_{23}	e_3	[e_1 , e_2 , e_3]	o_{15}	[1,1,1]	[0,0,1]	1	
	e_3	[e_1 , e_2 , e_3]	o_{20}	[1,1,1]	[1,1,1]	2	
t_{24}	∅	[e_1 , e_2 , e_3]	o_{15}	[1,1,1]	[0,1,1]	1	
	e_2	[e_1 , e_2 , e_3]	o_{16}	[1,1,1]	[1,1,1]	2	
t_{25}	e_1 , e_3	[e_1 , e_2 , e_3]	o_{20}	[1,1,1]	[1,1,1]	1	
	e_1 , e_3	[e_1 , e_2 , e_3]	o_{20}	[1,1,1]	[1,1,1]	2	
t_{26}	e_1 , e_2	[e_1 , e_2 , e_3]	o_{16}	[1,1,1]	[0,1,1]	1	
	e_1	[e_1 , e_2 , e_3]	o_{21}	[1,1,1]	[1,1,1]	2	
t_{27}	e_2	[e_1 , e_2 , e_3]	o_{16}	[1,1,1]	[0,1,1]	1	
	e_1 , e_3	[e_1 , e_2 , e_3]	o_{22}	[1,1,1]	[1,1,1]	2	
t_{28}	e_2	[e_1 , e_2 , e_3]	o_8	[1,1,1]	[0,1,1]	1	
	e_2	[e_1 , e_2 , e_3]	o_8	[1,1,1]	[1,1,1]	2	
t_{29}	e_2	[e_1 , e_2 , e_3]	o_8	[1,1,1]	[1,1,0]	2	
t_{30}	e_2	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,1]	2	
t_{31}	e_1 , e_2 , e_3	[e_1 , e_2 , e_3]	o_2	[0,1,1]	[1,1,1]	2	
t_{32}	e_2	[e_1 , e_2 , e_3]	o_3	[1,1,1]	[1,1,1]	2	
t_{33}	e_2	[e_1 , e_2 , e_3]	o_8	[1,1,1]	[1,1,1]	2	
t_{34}	e_2	[e_1 , e_2 , e_3]	o_2	[1,1,1]	[1,1,1]	2	
t_{35}	e_2	[e_1 , e_2 , e_3]	o_8	[1,1,1]	[1,1,1]	2	
t_{36}	e_2	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,1]	2	
t_{37}	e_2	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,1]	2	
t_{38}	e_2	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,1]	2	
t_{39}	e_2	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,1]	2	
t_{40}	e_1 , e_2 , e_3	[e_1 , e_2 , e_3]	o_{23}	[1,1,1]	[1,1,1]	2	
t_{41}	e_1	[e_1 , e_2 , e_3]	o_{19}	[1,1,1]	[1,1,1]	2	
t_{42}	e_1 , e_2 , e_3	[e_1 , e_2 , e_3]	o_{23}	[1,1,1]	[1,1,1]	2	

The responses of $VESA_1^1$, to the requests concerning the external knowledge EK_2 and compared with the responses of its human origin e_1 look as follows. Identical behavior (solutions or ratings) of ***VESA*** and human origin is marked red.

EK_2	solution of	
	$VESA_1^1$	e_1
t_{15}	o_{10}	o_3
t_{16}	o_9	o_3
t_{17}	o_1	o_{14}
t_{18}	o_6	o_{18}
t_{19}	o_{12}	o_{13}
t_{20}	o_1	o_{12}
t_{21}	o_{21}	o_{15}

EK_2	solution of	
	$VESA_1^1$	e_1
t_{22}	**o_{17}**	**o_{17}**
t_{23}	**o_{13}**	**o_{13}**
t_{24}	o_{16}	o_{19}
t_{25}	**o_{20}**	**o_{20}**
t_{26}	o_{16}	o_{21}
t_{27}	o_{18}	o_{22}
t_{28}	o_9	o_6

EK_2	solution	rating of		certainty of	
		$VESA_1^1$	e_1	$VESA_1^1$	e_1
t_{15}	o_3	**1**	**1**	1	1
	o_9	0	1	1	1
t_{16}	o_3	**1**	**1**	1	1
	o_9	0	1	1	1
	o_{11}	0	1	1	1
t_{17}	o_1	0	0	1	1
	o_{14}	**1**	**1**	1	1
	o_{18}	**0**	**0**	1	1
	o_{20}	0	1	1	1
t_{18}	o_1	**0**	**0**	1	1
	o_6	**0**	**0**	1	1
	o_{18}	1	0	1	1
	o_{20}	0	1	1	1
t_{19}	o_1	**0**	**0**	1	1
	o_{12}	0	1	1	1
	o_{13}	**1**	**1**	1	1
	o_{21}	**0**	**0**	1	1
t_{20}	o_1	**0**	**0**	1	1
	o_4	0	1	1	1
	o_{12}	**1**	**1**	1	1
	o_{21}	0	1	1	1
t_{21}	o_{15}	**1**	**1**	1	1
	o_{18}	**0**	**0**	1	1
	o_{21}	0	1	1	1
	o_{22}	0	1	1	1

EK_2	solution	rating of		certainty of	
		$VESA_1^1$	e_1	$VESA_1^1$	e_1
t_{22}	o_{14}	0	1	1	1
	o_{17}	**1**	**1**	1	1
	o_{18}	**0**	**0**	1	1
t_{23}	o_{13}	**1**	**1**	1	1
	o_{15}	**1**	**1**	0	0
	o_{16}	0	1	1	1
	o_{18}	**0**	**0**	1	1
	o_{20}	0	1	1	1
t_{24}	o_{15}	0	1	1	1
	o_{16}	0	1	1	1
	o_{19}	1	0	1	1
t_{25}	o_{15}	0	1	1	1
	o_{16}	0	1	1	1
	o_{20}	**1**	**1**	1	1
t_{26}	o_{16}	0	1	1	1
	o_{18}	0	1	1	1
	o_{20}	0	1	1	1
	o_{21}	**1**	**1**	1	1
t_{27}	o_{16}	0	1	1	1
	o_{18}	**0**	**0**	1	1
	o_{22}	**1**	**1**	1	1
t_{28}	o_6	**1**	**1**	1	1
	o_8	0	1	1	1
	o_{16}	**0**	**0**	1	1

Evaluation of the 2nd session:

- the number a_2 of cases from VKB^1, which were subject of the rating session and relate it to $|EK_2|$: $A_2 := a_2 / |EK_2|$

  - *There was just one case, for which VKB^1 had a solution which was not in the process anyway: o_{15} for test data t_{23} , i.e. $a_2 = 1$. This solution became worse marks in the 2nd session.* $A_2 := 1 / 14 \approx 0{,}071$

- the number b_2 of cases from VKB^1, which provided the optimal (best rated) solution and relate it to $|EK_2|$: $B_2 := b_2 / |EK_2|$
  - *This solution became worse marks in the 2nd session*, i.e. $b_2 = 0$: $B_2 := 0$
- the number c_2 of cases from VKB^1, for which a new solution has been introduced into VKB and relate it to $|EK_2|$: $C_2 := c_2 / |EK_2|$
  - *For $c_2 = 7$ of the 14 cases in EK_2 a new solution has been introduced in VKB^1 towards VKB^2:* $C_2 := 7 / 14 = 0{,}5$
- the number d_2 of solutions and ratings, which are identical responses of e_1 and $VESA_1$ and relate it to the number of required solutions and ratings: $D_2 := d_2 /$ ***number of expert responses altogether***
  - *For 3 (out of 14) cases $VESA_1$ provided the same solution as its human origin.*
  - *For 24 out of 49 rating requests $VESA_1$ provided the same rating as its human origin.*
  - *Thus, $d_2 = (3+24) = 27$:* $D_2 := 27 / 53 \approx 0{,}51$

7.3 3rd session

As a result
- of filling in the questionnaire shown in 10.5 by the three experts e_1, e_2, and e_3 and
- running the 28 test cases by the system

we received the following outputs o_k provided by the experts l_{ei} results to the test data t_j. Here, a new solution (different from all system's solutions) came up by expert e_1 : She provided the solution
- o_{25} Sauternes

to the test cases t_{37} , t_{38} , and t_{39}.

	e_1	e_2	e_3	system		e_1	e_2	e_3	system
t_1	o_6	o_{21}	o_4	o_{18}	t_{29}	o_3	o_8	o_4	o_{16}
t_2	o_7	o_{20}	o_2	o_2	t_{30}	o_4	o_9	o_2	o_{16}
t_3	o_8	o_{23}	o_3	o_2	t_{31}	o_8	o_2	o_4	o_{16}
t_4	o_2	o_{24}	o_2	o_{10}	t_{32}	o_9	o_3	o_3	o_{16}
t_5	o_8	o_3	o_1	o_{18}	t_{33}	o_9	o_8	o_3	o_{16}
t_6	o_9	o_9	o_2	o_2	t_{34}	o_{10}	o_2	o_4	o_{17}
t_7	o_3	o_8	o_3	o_2	t_{35}	o_{11}	o_8	o_5	o_{17}
t_8	o_4	o_4	o_2	o_{10}	t_{36}	o_{11}	o_9	o_1	o_{17}
t_9	o_6	o_{15}	o_5	o_1	t_{37}	o_{25}	o_9	o_{10}	o_{24}
t_{10}	o_2	o_3	o_2	o_1	t_{38}	o_{25}	o_9	o_2	o_{24}
t_{11}	o_{13}	o_8	o_3	o_1	t_{39}	o_{25}	o_9	o_3	o_{24}
t_{12}	o_5	o_9	o_4	o_1	t_{40}	o_{19}	o_{23}	o_{14}	o_{22}
t_{13}	o_8	o_9	o_5	o_3	t_{41}	o_{23}	o_{22}	o_{21}	o_{19}
t_{14}	o_9	o_4	o_2	o_3	t_{42}	o_{23}	o_{23}	o_{23}	o_{23}

This led to the questionnaire for the first rating session as listed in section 10.6. The rating session based on this questionnaire led to the following result. The own solution of each expert is marked red. The test data marked blue has been introduced from ***VKB***^{2}.

test data	solution	e_1		e_2		e_3	
		rating	certainty	rating	certainty	rating	certainty
t_1	o_4	1	1	0	1	**1**	**1**
t_1	o_6	**1**	**1**	0	1	1	1
t_1	o_{21}	0	1	**0**	**1**	0	0
t_1	o_{18}	0	1	1	1	1	0
t_2	o_2	1	1	0	1	**1**	**1**
t_2	o_7	**0**	**1**	0	1	1	1
t_2	o_{20}	0	1	**1**	**1**	0	1
t_3	o_2	1	1	0	1	1	1
t_3	o_3	1	1	0	1	**1**	**1**
t_3	o_8	**1**	**1**	0	1	1	1
t_3	o_{20}	0	1	0	1	0	1
t_3	o_{23}	1	1	**0**	**1**	0	1
t_4	o_2	**1**	**1**	0	1	**1**	**1**
t_4	o_{10}	1	1	0	1	1	1
t_4	o_{24}	0	1	**0**	**1**	0	0
t_5	o_1	0	1	1	1	**1**	**0**
t_5	o_3	1	1	**1**	**1**	1	1
t_5	o_8	**1**	**1**	1	0	1	1
t_5	o_{18}	0	1	0	1	0	1
t_6	o_2	1	1	1	0	**1**	**1**
t_6	o_9	**1**	**1**	**1**	**1**	1	1
t_7	o_2	1	1	1	0	1	1
t_7	o_3	**1**	**1**	1	0	**1**	**1**
t_7	o_8	1	1	**1**	**1**	1	1
t_8	o_2	1	1	0	1	**1**	**1**
t_8	o_3	1	1	0	1	1	1
t_8	o_4	**1**	**1**	**1**	**1**	1	1
t_8	o_{10}	1	1	1	1	1	1
t_9	o_1	1	1	0	1	1	1
t_9	o_5	1	1	0	1	**1**	**1**
t_9	o_6	**1**	**1**	0	1	1	1
t_9	o_{15}	1	1	**1**	**1**	0	1
t_{10}	o_1	1	1	0	1	1	0
t_{10}	o_2	**1**	**1**	0	1	**1**	**1**
t_{10}	o_3	1	1	**0**	**1**	1	1
t_{11}	o_1	1	1	0	1	0	0
t_{11}	o_3	1	1	0	1	**1**	**1**
t_{11}	o_8	1	1	**0**	**1**	1	1
t_{11}	o_{13}	**0**	**1**	1	0	0	1

test data	solution	e_1		e_2		e_3	
		rating	certainty	rating	certainty	rating	certainty
t_{12}	o_1	1	1	0	1	1	0
t_{12}	o_4	1	1	1	1	**1**	**0**
t_{12}	o_5	**1**	**1**	0	1	1	0
t_{12}	o_9	1	1	**1**	**1**	1	1
t_{13}	o_3	1	1	1	1	1	1
t_{13}	o_5	0	1	1	1	**1**	**1**
t_{13}	o_8	**1**	**1**	0	1	1	1
t_{13}	o_9	1	1	**0**	**1**	1	1
t_{14}	o_2	1	1	1	0	**1**	**1**
t_{14}	o_3	1	1	1	0	1	1
t_{14}	o_4	1	1	**1**	**1**	1	0
t_{14}	o_9	**1**	**1**	1	1	1	1
t_{29}	o_3	**1**	**1**	0	1	1	1
t_{29}	o_4	1	1	1	1	**1**	**1**
t_{29}	o_8	1	1	**1**	**1**	1	1
t_{29}	o_{16}	1	0	0	1	1	1
t_{30}	o_2	1	1	0	1	**1**	**0**
t_{30}	o_4	**1**	**1**	1	1	1	1
t_{30}	o_9	1	1	**1**	**1**	1	0
t_{30}	o_{16}	1	0	0	1	1	1
t_{31}	o_2	1	1	**0**	**1**	1	0
t_{31}	o_4	1	1	1	1	**1**	**1**
t_{31}	o_8	**1**	**1**	1	1	1	1
t_{31}	o_{16}	1	0	0	1	1	1
t_{32}	o_3	1	1	**1**	**0**	**1**	**1**
t_{32}	o_9	**1**	**1**	1	1	1	0
t_{32}	o_{16}	1	1	0	1	1	1
t_{33}	o_3	1	1	0	1	**1**	**1**
t_{33}	o_8	1	1	**1**	**1**	1	1
t_{33}	o_9	**1**	**1**	1	1	1	0
t_{33}	o_{16}	1	1	0	1	1	1
t_{34}	o_2	1	1	**1**	**1**	0	1
t_{34}	o_4	1	1	1	1	**1**	**1**
t_{34}	o_{10}	**1**	**1**	1	1	1	0
t_{34}	o_{17}	1	0	0	1	1	1
t_{35}	o_5	0	1	0	1	**1**	**1**
t_{35}	o_8	1	1	**1**	**1**	1	1
t_{35}	o_{11}	**1**	**1**	1	1	1	0
t_{35}	o_{17}	1	0	0	1	0	1
t_{36}	o_1	0	1	0	1	**1**	**1**
t_{36}	o_9	1	1	**1**	**1**	1	0
t_{36}	o_{11}	**1**	**1**	1	1	1	0
t_{36}	o_{17}	1	0	0	1	1	1
t_{37}	o_9	1	1	**1**	**1**	1	1
t_{37}	o_{10}	1	1	1	1	**1**	**1**

test data	solution	e_1		e_2		e_3	
		rating	certainty	rating	certainty	rating	certainty
t_{37}	o_{24}	0	1	0	1	0	0
t_{37}	o_{25}	**0**	**0**	0	0	0	1
t_{38}	o_2	1	1	1	0	**1**	**1**
t_{38}	o_9	1	1	**1**	**1**	1	1
t_{38}	o_{24}	0	1	0	1	0	0
t_{38}	o_{25}	**1**	**1**	0	0	0	1
t_{39}	o_3	1	1	1	0	**1**	**1**
t_{39}	o_9	1	1	**1**	**1**	1	1
t_{39}	o_{24}	0	1	0	1	0	0
t_{39}	o_{25}	**1**	**1**	0	0	0	1
t_{40}	o_{14}	0	1	0	1	**1**	**1**
t_{40}	o_{19}	**1**	**1**	1	0	1	1
t_{40}	o_{22}	0	1	1	0	1	0
t_{40}	o_{23}	1	1	**1**	**1**	0	1
t_{41}	o_{19}	1	1	1	1	1	1
t_{41}	o_{21}	0	1	1	0	**1**	**1**
t_{41}	o_{22}	0	1	**0**	**0**	1	1
t_{41}	o_{23}	**1**	**1**	1	0	0	1
t_{42}	o_{23}	**1**	**1**	**1**	**1**	**1**	**1**

The competences of each expert per test case following the approach of (Knauf 2000) and the optimal solution that gained the best certain rating (weighted by the experts' competences each) of the expert panel are as follows (represented as fractions for exactness). In case concurrent solutions enjoyed optimality, the sum of the competences of the (also uncertain) supporters decides. If this is the same, the sum of the competences of solvers decides.
In case this is still the same number, we used here the overall competence (average over all test cases). The solution provided by the most competent expert over all test cases enjoys optimality.
If this is still not a KO criterion, the systems solution is preferred.

test data	$cpt(e_1, t_j)$	$cpt(e_2, t_j)$	$cpt(e_3, t_j)$	sol_{Kj}^{opt}
t_1	5 / 6	1 / 2	7 / 9	o_4
t_2	2 / 3	2 / 3	5 / 6	o_2
t_3	5 / 6	2 / 3	5 / 6	o_8
t_4	5 / 6	1 / 2	3 / 4	o_2
t_5	1	17 / 18	2 / 3	o_3
t_6	1	5 / 6	1	o_9
t_7	1	5 / 6	1	o_8
t_8	1	1	5 / 6	o_4
t_9	5 / 6	5 / 6	5 / 6	o_6
t_{10}	5 / 6	5 / 6	3 / 4	o_2
t_{11}	1 / 2	7 / 9	7 / 9	o_8
t_{12}	2 / 3	1	7 / 9	o_9
t_{13}	5 / 6	5 / 6	5 / 6	o_3
t_{14}	1	8 / 9	17 / 18	o_9

test data	$cpt\ (e_1, t_j)$	$cpt\ (e_2, t_j)$	$cpt\ (e_3, t_j)$	sol_{Kj}^{opt}
t_{29}	7 / 9	1	1	o_8
t_{30}	17 / 18	1	7 / 9	o_4
t_{31}	17 / 18	5 / 6	17 / 18	o_8
t_{32}	1	5 / 6	11 / 12	o_3
t_{33}	1	1	7 / 9	o_8
t_{34}	17 / 18	5 / 6	17 / 18	o_4
t_{35}	17 / 18	1	11 / 18	o_8
t_{36}	17 / 18	1	5 / 9	o_9
t_{37}	2 / 3	17 / 18	17 / 18	o_9
t_{38}	2 / 3	8 / 9	17 / 18	o_9
t_{39}	2 / 3	8 / 9	17 / 18	o_9
t_{40}	1	11 / 18	11 / 18	o_{19}
t_{41}	2 / 3	7 / 18	2 / 3	o_{19}
t_{42}	1	1	1	o_{23}
overall competence	24	23,33	23,25	

The resulting VKB^3 is

t_j	E_K	E_I	sol_{Kj}^{opt}	r_{ijk}	c_{ijk}	τ_s	D_C
t_1	e_1, e_3	$[e_1, e_2, e_3]$	o_6	[1,0,1]	[0,1,1]	1	
	e_3	$[e_1, e_2, e_3]$	o_4	[1,0,1]	[1,1,1]	3	
t_2	e_1, e_3	$[e_1, e_2, e_3]$	o_7	[0,0,1]	[0,0,1]	1	
	e_3	$[e_1, e_2, e_3]$	o_2	[1,0,1]	[1,1,1]	3	
t_3	e_2	$[e_1, e_2, e_3]$	o_{20}	[0,1,0]	[0,1,1]	1	
	e_1	$[e_1, e_2, e_3]$	o_8	[1,0,1]	[1,1,1]	3	
t_4	e_1, e_3	$[e_1, e_2, e_3]$	o_2	[0,1,1]	[0,1,1]	1	
	e_1, e_3	$[e_1, e_2, e_3]$	o_2	[1,0,1]	[1,1,1]	3	
t_5	e_1, e_3	$[e_1, e_2, e_3]$	o_8	[1,1,1]	[1,0,1]	1	
	e_2	$[e_1, e_2, e_3]$	o_3	[1,1,1]	[1,1,1]	3	
t_6	e_1, e_2, e_3	$[e_1, e_2, e_3]$	o_9	[1,1,1]	[0,1,1]	1	
	e_1, e_2	$[e_1, e_2, e_3]$	o_9	[1,1,1]	[1,1,1]	3	
t_7	e_1, e_2, e_3	$[e_1, e_2, e_3]$	o_3	[1,1,1]	[0,1,1]	1	
	e_2	$[e_1, e_2, e_3]$	o_8	[1,1,1]	[1,1,1]	3	
t_8	e_2, e_3	$[e_1, e_2, e_3]$	o_3	[1,1,1]	[0,1,1]	1	
	e_1, e_2	$[e_1, e_2, e_3]$	o_4	[1,1,1]	[1,1,1]	3	
t_9	e_1	$[e_1, e_2, e_3]$	o_1	[1,1,1]	[0,1,1]	1	
	e_1	$[e_1, e_2, e_3]$	o_6	[1,0,1]	[1,1,1]	3	
t_{10}	$\varnothing$	$[e_1, e_2, e_3]$	o_1	[1,1,1]	[0,1,1]	1	
	e_1, e_3	$[e_1, e_2, e_3]$	o_2	[1,0,1]	[1,1,1]	3	
t_{11}	$\varnothing$	$[e_1, e_2, e_3]$	o_1	[1,1,1]	[1,1,1]	1	
	e_2	$[e_1, e_2, e_3]$	o_8	[1,0,1]	[1,1,1]	3	
t_{12}	e_3	$[e_1, e_2, e_3]$	o_1	[1,1,1]	[1,1,1]	1	
	e_2	$[e_1, e_2, e_3]$	o_9	[1,1,1]	[1,1,1]	3	
t_{13}	e_2, e_3	$[e_1, e_2, e_3]$	o_9	[1,1,1]	[0,1,1]	1	
	$\varnothing$	$[e_1, e_2, e_3]$	o_3	[1,1,1]	[1,1,1]	3	

t_j	E_K	E_I	sol_{Kj}^{opt}	r_{ijk}	c_{ijk}	τ_s	D_C
t_{14}	e_3	[e_1 , e_2 , e_3]	o_3	[0,1,1]	[0,1,1]	1	
	e_1	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,1]	3	
t_{15}	e_3	[e_1 , e_2 , e_3]	o_3	[0,1,1]	[0,1,1]	1	
	e_1 , e_3	[e_1 , e_2 , e_3]	o_3	[1,1,1]	[1,1,1]	2	
t_{16}	e_2	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,1]	1	
	e_1	[e_1 , e_2 , e_3]	o_3	[1,1,1]	[1,1,1]	2	
t_{17}	e_3	[e_1 , e_2 , e_3]	o_{20}	[1,1,1]	[0,1,1]	1	
	e_3	[e_1 , e_2 , e_3]	o_{20}	[1,1,1]	[1,1,0]	2	
t_{18}	e_1 , e_3	[e_1 , e_2 , e_3]	o_6	[0,1,1]	[0,1,1]	1	
	e_3	[e_1 , e_2 , e_3]	o_6	[0,1,1]	[1,1,1]	2	
t_{19}	e_1 , e_3	[e_1 , e_2 , e_3]	o_{12}	[1,1,1]	[0,1,1]	1	
	e_3	[e_1 , e_2 , e_3]	o_{12}	[1,0,1]	[1,1,1]	2	
t_{20}	e_2	[e_1 , e_2 , e_3]	o_{21}	[1,1,0]	[0,1,1]	1	
	e_3	[e_1 , e_2 , e_3]	o_4	[1,0,1]	[1,1,1]	2	
t_{21}	e_2	[e_1 , e_2 , e_3]	o_{21}	[1,1,1]	[1,1,1]	1	
	e_1	[e_1 , e_2 , e_3]	o_{15}	[1,1,1]	[1,1,1]	2	
t_{22}	e_1 , e_3	[e_1 , e_2 , e_3]	o_{17}	[1,1,1]	[1,1,1]	1	
	e_1	[e_1 , e_2 , e_3]	o_{17}	[1,1,1]	[1,1,1]	2	
t_{23}	e_3	[e_1 , e_2 , e_3]	o_{15}	[1,1,1]	[0,0,1]	1	
	e_3	[e_1 , e_2 , e_3]	o_{20}	[1,1,1]	[1,1,1]	2	
t_{24}	∅	[e_1 , e_2 , e_3]	o_{15}	[1,1,1]	[0,1,1]	1	
	e_2	[e_1 , e_2 , e_3]	o_{16}	[1,1,1]	[1,1,1]	2	
t_{25}	e_1 , e_3	[e_1 , e_2 , e_3]	o_{20}	[1,1,1]	[1,1,1]	1	
	e_1 , e_3	[e_1 , e_2 , e_3]	o_{20}	[1,1,1]	[1,1,1]	2	
t_{26}	e_1 , e_2	[e_1 , e_2 , e_3]	o_{16}	[1,1,1]	[0,1,1]	1	
	e_1	[e_1 , e_2 , e_3]	o_{21}	[1,1,1]	[1,1,1]	2	
t_{27}	e_2	[e_1 , e_2 , e_3]	o_{16}	[1,1,1]	[0,1,1]	1	
	e_1 , e_3	[e_1 , e_2 , e_3]	o_{22}	[1,1,1]	[1,1,1]	2	
t_{28}	e_2	[e_1 , e_2 , e_3]	o_8	[1,1,1]	[0,1,1]	1	
	e_2	[e_1 , e_2 , e_3]	o_8	[1,1,1]	[1,1,1]	2	
t_{29}	e_2	[e_1 , e_2 , e_3]	o_8	[1,1,1]	[1,1,0]	2	
	e_2	[e_1 , e_2 , e_3]	o_8	[1,1,1]	[1,1,1]	3	
t_{30}	e_2	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,1]	2	
	e_1	[e_1 , e_2 , e_3]	o_4	[1,1,1]	[1,1,1]	3	
t_{31}	e_1 , e_2 , e_3	[e_1 , e_2 , e_3]	o_2	[0,1,1]	[1,1,1]	2	
	e_1	[e_1 , e_2 , e_3]	o_8	[1,1,1]	[1,1,1]	3	
t_{32}	e_2	[e_1 , e_2 , e_3]	o_3	[1,1,1]	[1,1,1]	2	
	e_2 , e_3	[e_1 , e_2 , e_3]	o_3	[1,1,1]	[1,0,1]	3	
t_{33}	e_2	[e_1 , e_2 , e_3]	o_8	[1,1,1]	[1,1,1]	2	
	e_2	[e_1 , e_2 , e_3]	o_8	[1,1,1]	[1,1,1]	3	
t_{34}	e_2	[e_1 , e_2 , e_3]	o_2	[1,1,1]	[1,1,1]	2	
	e_3	[e_1 , e_2 , e_3]	o_4	[1,1,1]	[1,1,1]	3	
t_{35}	e_2	[e_1 , e_2 , e_3]	o_8	[1,1,1]	[1,1,1]	2	
	e_2	[e_1 , e_2 , e_3]	o_8	[1,1,1]	[1,1,1]	3	
t_{36}	e_2	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,1]	2	

t_j	E_K	E_I	sol_{Kj}^{opt}	r_{ijk}	c_{ijk}	τ_s	D_C
	e_2	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,0]	3	
t_{37}	e_2	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,1]	2	
	e_2	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,1]	3	
t_{38}	e_2	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,1]	2	
	e_2	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,1]	3	
t_{39}	e_2	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,1]	2	
	e_2	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,1]	3	
t_{40}	e_1 , e_2 , e_3	[e_1 , e_2 , e_3]	o_{23}	[1,1,1]	[1,1,1]	2	
	e_1	[e_1 , e_2 , e_3]	o_{19}	[1,1,1]	[1,0,1]	3	
t_{41}	e_1	[e_1 , e_2 , e_3]	o_{19}	[1,1,1]	[1,1,1]	2	
	∅	[e_1 , e_2 , e_3]	o_{19}	[1,1,1]	[1,1,1]	3	
t_{42}	e_1 , e_2 , e_3	[e_1 , e_2 , e_3]	o_{23}	[1,1,1]	[1,1,1]	2	
	e_1 , e_2 , e_3	[e_1 , e_2 , e_3]	o_{23}	[1,1,1]	[1,1,1]	3	

Unfortunately, in this setting of the experiment there is never a need to compute a "most similar expert", because each expert solved or rated the considered case in former sessions.
To validate this approach nevertheless, we include the calculation of a most similar expert for e_2 and compare his/her replies with the one of $VESA_2^2$ as well:

		# of identical solutions	# of identical ratings	Σ per session	Σ at all
e_1	1st session	3	62	65	208
	2nd session	3	81	84	
	3rd session	0	59	59	
e_3	1st session	5	62	67	198
	2nd session	4	67	71	
	3rd session	0	60	60	

Thus, e_3 is the most similar expert to e_2.

The responses of $VESA_2^2$, to the requests concerning the external knowledge EK_3 and compared with the responses of its human origin e_2 look as follows. Identical behavior (solutions or ratings) of *VESA* and human origin is marked red.

EK_3	solution of			EK_3	solution of			EK_3	solution of			EK_3	solution of		
	$VESA_2^2$	e_2	e_3		$VESA_2^2$	e_2	e_3		$VESA_2^2$	e_2	e_3		$VESA_2^2$	e_2	e_3
t_1	o_{18}	o_{21}	o_4	t_8	**o_4**	**o_4**	o_2	t_{29}	**o_8**	**o_8**	o_4	t_{36}	**o_9**	**o_9**	o_1
t_2	o_{21}	o_{20}	o_2	t_9	o_{20}	o_{15}	o_5	t_{30}	**o_9**	**o_9**	o_2	t_{37}	**o_9**	**o_9**	o_{10}
t_3	o_{20}	o_{23}	o_3	t_{10}	o_{12}	o_3	o_2	t_{31}	**o_2**	**o_2**	o_4	t_{38}	**o_9**	**o_9**	o_2
t_4	o_1	o_{24}	o_2	t_{11}	o_4	o_8	o_3	t_{32}	**o_8**	**o_3**	o_3	t_{39}	**o_9**	**o_9**	o_3
t_5	**o_3**	**o_3**	o_1	t_{12}	o_4	o_9	o_4	t_{33}	**o_8**	**o_8**	o_3	t_{40}	**o_{23}**	**o_{23}**	o_{14}
t_6	**o_9**	**o_9**	o_2	t_{13}	**o_9**	**o_9**	o_5	t_{34}	**o_2**	**o_2**	o_4	t_{41}	o_{19}	o_{22}	o_{21}
t_7	o_3	o_8	o_3	t_{14}	o_9	o_4	o_2	t_{35}	**o_8**	**o_8**	o_5	t_{42}	**o_{23}**	**o_{23}**	o_{23}

EK_3	solution	rating of			certainty of			EK_3	solution	rating of			certainty of		
		$VESA_2^2$	e_2	e_3	$VESA_2^2$	e_2	e_3			$VESA_2^2$	e_2	e_3	$VESA_2^2$	e_2	e_3
t_1	o_4	0	0	1	1	1	1	t_{29}	o_3	0	0	1	1	1	1
	o_6	0	0	1	1	1	1		o_4	0	1	1	1	1	1
	o_{21}	0	0	0	1	0	0		o_8	1	1	1	1	1	1
	o_{18}	1	1	1	1	0	0		o_{16}	0	0	1	1	1	1
t_2	o_2	0	0	1	1	1	1	t_{30}	o_2	0	0	1	1	0	0
	o_7	0	0	1	0	1	1		o_4	0	1	1	1	1	1
	o_{20}	0	1	0	1	1	1		o_9	1	1	1	1	0	0
t_3	o_2	0	0	1	1	1	1		o_{16}	0	0	1	1	1	1
	o_3	0	0	1	1	1	1	t_{31}	o_2	1	0	1	1	0	0
	o_8	0	0	1	1	1	1		o_4	0	1	1	1	1	1
	o_{20}	1	0	0	1	1	1		o_8	0	1	1	1	1	1
	o_{23}	0	0	0	1	1	1		o_{16}	0	0	1	1	1	1
t_4	o_2	1	0	1	1	1	1	t_{32}	o_3	1	1	1	1	1	1
	o_{10}	0	0	1	0	1	1		o_9	0	1	1	1	0	0
	o_{24}	0	0	0	1	0	0		o_{16}	0	0	1	1	1	1
t_5	o_1	1	1	1	0	0	0	t_{33}	o_3	1	0	1	1	1	1
	o_3	1	1	1	0	1	1		o_8	1	1	1	1	1	1
	o_8	1	1	1	0	1	1		o_9	0	1	1	1	0	0
	o_{18}	0	0	0	0	1	1		o_{16}	0	0	1	1	1	1
t_6	o_2	1	1	1	0	1	1	t_{34}	o_2	1	1	0	1	1	1
	o_9	1	1	1	1	1	1		o_4	0	1	1	1	1	1
t_7	o_2	1	1	1	1	1	1		o_{10}	0	1	1	1	0	0
	o_3	1	1	1	1	1	1		o_{17}	0	0	1	1	1	1
	o_8	0	1	1	1	1	1	t_{35}	o_5	0	0	1	1	1	1
t_8	o_2	0	0	1	1	1	1		o_8	1	1	1	1	1	1
	o_3	1	0	1	1	1	1		o_{11}	0	1	1	1	0	0
	o_4	1	1	1	1	1	1		o_{17}	0	0	0	1	1	1
	o_{10}	0	1	1	0	1	1	t_{36}	o_1	0	0	1	1	1	1
t_9	o_1	1	0	1	1	1	1		o_9	1	1	1	1	0	0
	o_5	0	0	1	1	1	1		o_{11}	1	1	1	1	0	0
	o_6	0	0	1	1	1	1		o_{17}	0	0	1	1	1	1
	o_{15}	0	1	0	1	1	1	t_{37}	o_9	1	1	1	1	1	1
t_{10}	o_1	1	0	1	1	0	0		o_{10}	1	1	1	1	1	1
	o_2	1	0	1	1	1	1		o_{24}	0	0	0	1	0	0
	o_3	0	0	1	1	1	1		o_{25}	0	0	0	1	1	1
t_{11}	o_1	1	0	0	1	0	0	t_{38}	o_2	0	1	1	1	1	1
	o_3	0	0	1	1	1	1		o_9	1	1	1	1	1	1
	o_8	0	0	1	1	1	1		o_{24}	0	0	0	1	0	0
	o_{13}	0	1	0	1	1	1		o_{25}	0	0	0	1	1	1
t_{12}	o_1	1	0	1	1	0	0	t_{39}	o_3	0	1	1	1	1	1
	o_4	1	1	1	0	0	0		o_9	1	1	1	1	1	1
	o_5	1	0	1	1	0	0		o_{24}	0	0	0	1	0	0
	o_9	0	1	1	1	1	1		o_{25}	0	0	0	1	1	1
t_{13}	o_3	1	1	1	1	1	1	t_{40}	o_{14}	0	0	1	1	1	1
	o_5	0	1	1	1	1	1		o_{19}	0	1	1	1	1	1
	o_8	1	0	1	1	1	1		o_{22}	0	1	1	1	0	0

	o_9	1	0	1	1	1	1
t_{14}	o_2	0	1	1	1	1	1
	o_3	**1**	**1**	1	1	1	1
	o_4	0	1	1	1	0	0
	o_9	**1**	**1**	1	1	1	1

	o_{23}	**1**	**1**	0	1	1	1
t_{41}	o_{19}	**1**	**1**	1	1	1	1
	o_{21}	0	1	1	1	1	1
	o_{22}	**0**	**0**	1	1	1	1
	o_{23}	0	1	0	1	1	1
t_{42}	o_{23}	**1**	**1**	1	1	1	1

Evaluation of the 3rd session:

- the number a_3 of cases from VKB^2, which were subject of the rating session and relate it to $|EK_3|$: $A_3 := a_3 / |EK_3|$
  - *There were two cases, for which VKB^2 had a solution which was not in the process anyway: o_3 for test data t_8 and o_{20} for test data t_3 , i.e. $a_3 = 2$. This solution became worse marks in the 2nd session.* $A_3 := 2 / 28 \approx 0{,}071$
- the number b_3 of cases from VKB^2, which provided the optimal (best rated) solution and relate it to $|EK_3|$: $B_3 := b_3 / |EK_3|$
  - *Both new solutions did not become optimal in the rating process,* i.e. $b_3 = 0$: $B_3 := 0$
- the number c_3 of cases from VKB^2, for which a new solution has been introduced into *VKB* and relate it to $|EK_3|$: $C_3 := c_3 / |EK_3|$
  - *For $c_3 = 16$ of the 28 cases in EK_3 a new solution has been introduced in VKB^2 towards VKB^3:* $C_3 := 16 / 28 = 0{,}57$
- the number d_3 of solutions and ratings, which are identical responses of e_2 and $VESA_2$ and relate it to the number of required solutions and ratings: $D_3 := d_3 /$ ***number of expert responses altogether***
  - *For 17 (out of 28) cases $VESA_2$ provided the same solution as its human origin.*
  - *For 61 (out of 98) rating requests $VESA_2$ provided the same rating as its human origin.*
  - *Thus, $d_3 = (17+61) = 79$:* $D_3 := 79 / 126 \approx 0{,}63$

7.4 4th session

As a result
- of filling in the questionnaire shown in 10.7 by the three experts e_1, e_2, and e_3 and
- running the 28 test cases by the system

we received the following outputs o_k provided by the experts e_i results to the test data t_j. New solutions (different from all solutions so far) did not occur.

	e_1	e_2	e_3	system		e_1	e_2	e_3	system
t_1	o_6	o_{17}	o_2	o_{18}	t_{22}	o_{17}	o_{14}	o_{20}	o_{14}

	e_1	e_2	e_3	system		e_1	e_2	e_3	system
t_2	o_7	o_{15}	o_5	o_2	t_{23}	o_{13}	o_{14}	o_{16}	o_{18}
t_4	o_2	o_{20}	o_2	o_{10}	t_{25}	o_{20}	o_{17}	o_{22}	o_{15}
t_5	o_8	o_5	o_5	o_{18}	t_{26}	o_{16}	o_{16}	o_{21}	o_{18}
t_7	o_3	o_6	o_{10}	o_2	t_{28}	o_9	o_2	o_6	o_{16}
t_8	o_4	o_9	o_4	o_{10}	t_{29}	o_3	o_3	o_{12}	o_{16}
t_{10}	o_2	o_1	o_9	o_1	t_{31}	o_8	o_9	o_7	o_{16}
t_{11}	o_{13}	o_3	o_{11}	o_1	t_{32}	o_9	o_8	o_4	o_{16}
t_{13}	o_8	o_{12}	o_{10}	o_3	t_{34}	o_{10}	o_9	o_8	o_{17}
t_{14}	o_9	o_9	o_9	o_3	t_{35}	o_{11}	o_9	o_5	o_{17}
t_{16}	o_{11}	o_8	o_7	o_{11}	t_{37}	o_{25}	o_8	o_{10}	o_{24}
t_{17}	o_{14}	o_{14}	o_6	o_1	t_{38}	o_{25}	o_8	o_4	o_{24}
t_{19}	o_{17}	o_{17}	o_3	o_1	t_{40}	o_{19}	o_{18}	o_{19}	o_{22}
t_{20}	o_{12}	o_{16}	o_1	o_1	t_{41}	o_{23}	o_{19}	o_{14}	o_{19}

This led to the questionnaire for the first rating session as listed in section 10.8. The rating session based on this questionnaire led to the following result. The own solution of each expert is marked red. The test data marked blue has been introduced from ***VKB***[3].

test data	solution	e_1		e_2		e_3	
		rating	certainty	rating	certainty	rating	certainty
t_1	o_2	1	1	0	1	**0**	**0**
t_1	o_4	1	1	0	1	1	1
t_1	o_6	**1**	**1**	0	1	1	1
t_1	o_{17}	0	1	**1**	**1**	0	1
t_1	o_{18}	0	1	0	0	0	1
t_2	o_2	1	1	0	1	1	1
t_2	o_5	1	1	0	1	**1**	**0**
t_2	o_7	**0**	**1**	0	0	1	1
t_2	o_{15}	0	1	**1**	**1**	0	1
t_4	o_2	**1**	**1**	1	1	**1**	**1**
t_4	o_{10}	1	1	0	1	1	1
t_4	o_{20}	0	1	**1**	**1**	0	1
t_5	o_3	1	1	1	1	1	1
t_5	o_5	1	1	**1**	**1**	**1**	**1**
t_5	o_8	**1**	**1**	1	1	1	1
t_5	o_{18}	0	1	0	1	0	1
t_7	o_2	1	1	1	1	1	1
t_7	o_3	**1**	**1**	0	1	1	1
t_7	o_6	1	1	**1**	**1**	1	0
t_7	o_8	1	1	1	1	1	0
t_7	o_{10}	1	1	1	0	**1**	**0**
t_8	o_3	1	1	0	1	1	1
t_8	o_4	**1**	**1**	1	1	**1**	**1**
t_8	o_9	1	1	**1**	**1**	1	1
t_8	o_{10}	1	1	1	0	1	1
t_{10}	o_1	1	1	**1**	**1**	1	0

test data	solution	e_1		e_2		e_3	
		rating	certainty	rating	certainty	rating	certainty
t_{10}	o_2	1	1	1	1	1	1
t_{10}	o_9	1	1	1	1	1	1
t_{11}	o_1	1	1	1	1	1	0
t_{11}	o_3	1	1	1	1	1	1
t_{11}	o_8	1	1	1	1	1	1
t_{11}	o_{11}	1	1	1	1	1	1
t_{11}	o_{13}	0	1	0	0	0	1
t_{13}	o_3	1	1	1	1	1	0
t_{13}	o_8	1	1	1	1	1	1
t_{13}	o_9	1	1	1	1	1	1
t_{13}	o_{10}	1	1	1	1	1	1
t_{13}	o_{12}	0	1	1	1	1	1
t_{14}	o_3	1	1	1	1	1	1
t_{14}	o_9	1	1	1	1	1	1
t_{16}	o_3	1	1	1	1	1	1
t_{16}	o_7	0	1	0	0	1	1
t_{16}	o_8	1	1	1	1	1	1
t_{16}	o_9	1	1	1	1	1	1
t_{16}	o_{11}	1	1	1	0	1	1
t_{17}	o_1	0	1	0	1	1	1
t_{17}	o_6	0	1	0	1	1	1
t_{17}	o_{14}	1	1	1	1	0	1
t_{17}	o_{20}	1	1	1	1	0	1
t_{19}	o_1	0	1	0	1	1	1
t_{19}	o_3	0	1	0	1	1	1
t_{19}	o_{12}	0	1	0	1	1	1
t_{19}	o_{17}	1	1	1	1	1	0
t_{20}	o_1	1	1	0	1	1	1
t_{20}	o_4	1	1	0	1	1	1
t_{20}	o_{12}	1	1	0	1	1	1
t_{20}	o_{16}	0	1	1	1	0	1
t_{20}	o_{21}	0	1	1	1	0	1
t_{22}	o_{14}	1	1	1	1	1	0
t_{22}	o_{17}	1	1	0	1	1	1
t_{22}	o_{20}	1	1	1	1	1	1
t_{23}	o_{13}	1	1	0	0	0	1
t_{23}	o_{14}	1	1	1	1	1	0
t_{23}	o_{15}	1	1	0	1	1	1
t_{23}	o_{16}	1	1	0	0	1	1
t_{23}	o_{18}	0	1	0	1	0	1
t_{23}	o_{20}	1	1	1	1	1	1
t_{25}	o_{15}	1	1	1	1	1	1
t_{25}	o_{17}	1	1	1	1	1	1
t_{25}	o_{20}	1	1	1	1	1	1
t_{25}	o_{22}	1	1	0	0	1	1

test data	solution	e_1		e_2		e_3	
		rating	certainty	rating	certainty	rating	certainty
t_{26}	o_{16}	1	1	1	0	1	1
t_{26}	o_{18}	0	1	0	1	0	1
t_{26}	o_{21}	1	1	1	1	1	1
t_{28}	o_2	1	1	1	1	1	0
t_{28}	o_6	1	1	1	1	1	1
t_{28}	o_8	1	1	1	1	1	1
t_{28}	o_9	1	1	1	1	1	1
t_{28}	o_{16}	1	0	0	1	0	1
t_{29}	o_3	1	1	1	1	1	0
t_{29}	o_8	1	1	1	1	1	1
t_{29}	o_{12}	1	1	1	1	1	1
t_{29}	o_{16}	1	0	0	1	0	1
t_{31}	o_2	1	1	1	1	1	0
t_{31}	o_7	0	1	0	0	1	1
t_{31}	o_8	1	1	1	1	1	1
t_{31}	o_9	1	1	1	1	1	1
t_{31}	o_{16}	1	1	0	1	0	1
t_{32}	o_3	1	1	1	1	1	1
t_{32}	o_4	1	1	1	1	1	1
t_{32}	o_8	1	1	1	1	1	1
t_{32}	o_9	1	1	1	1	1	1
t_{32}	o_{16}	1	1	0	1	0	1
t_{34}	o_2	1	0	1	1	1	0
t_{34}	o_4	1	1	1	1	1	1
t_{34}	o_8	1	1	1	1	1	1
t_{34}	o_9	1	1	1	1	1	1
t_{34}	o_{10}	1	1	1	1	1	1
t_{34}	o_{17}	1	1	0	1	0	1
t_{35}	o_5	1	1	0	0	1	1
t_{35}	o_8	1	1	1	1	1	1
t_{35}	o_9	1	1	1	1	1	1
t_{35}	o_{11}	1	1	1	0	1	1
t_{35}	o_{17}	1	0	0	1	0	1
t_{37}	o_8	1	1	1	1	1	1
t_{37}	o_9	1	1	1	1	1	1
t_{37}	o_{10}	1	1	1	0	1	1
t_{37}	o_{24}	0	1	0	1	*norating*	0
t_{37}	o_{25}	1	1	0	0	0	1
t_{38}	o_4	1	1	1	1	1	1
t_{38}	o_8	1	1	1	1	1	1
t_{38}	o_9	1	1	1	1	1	1
t_{38}	o_{24}	0	1	0	1	*norating*	0
t_{38}	o_{25}	1	1	0	0	0	1
t_{40}	o_{18}	1	1	1	0	1	1
t_{40}	o_{19}	1	1	1	0	1	1

test data	solution	e_1		e_2		e_3	
		rating	certainty	rating	certainty	rating	certainty
t_{40}	o_{22}	0	1	0	1	1	1
t_{40}	o_{23}	1	1	1	0	1	1
t_{41}	o_{14}	0	1	1	0	**1**	**1**
t_{41}	o_{19}	1	1	**1**	**0**	1	1
t_{41}	o_{23}	**1**	**1**	1	0	1	1

The competences of each expert per test case following the approach of (Knauf 2000) and the optimal solution that gained the best certain rating (weighted by the experts' competences each) of the expert panel are as follows (represented as fractions for exactness). In case concurrent solutions enjoyed optimality, the sum of the competences of the (also uncertain) supporters decides. If this is the same, the sum of the competences of solvers decides.
In case this is still the same number, we used here the overall competence (average over all test cases). The solution provided by the most competent expert over all test cases enjoys optimality.
If this is still not a KO criterion, the systems solution is preferred.

test data	$cpt(e_1, t_j)$	$cpt(e_2, t_j)$	$cpt(e_3, t_j)$	sol_{Ki}^{opt}
t_1	5 / 6	5 / 8	1 / 2	o_{17}
t_2	5 / 6	11 / 18	2 / 3	o_2
t_4	1	2 / 3	1	o_2
t_5	1	1	1	o_8
t_7	5 / 6	23 / 24	3 / 4	o_2
t_8	1	17 / 18	1	o_4
t_{10}	1	1	1	o_2
t_{11}	1 / 2	23 / 24	11 / 12	o_3
t_{13}	1	5 / 6	23 / 24	o_8
t_{14}	1	1	23 / 24	o_9
t_{16}	1	11 / 12	2 / 3	o_8
t_{17}	5 / 6	5 / 6	2 / 3	o_{14}
t_{19}	1	1	2 / 3	o_{17}
t_{20}	5 / 6	2 / 3	7 / 9	o_{12}
t_{22}	5 / 6	1	1	o_{20}
t_{23}	2 / 3	14 / 15	11 / 12	o_{20}
t_{25}	1	17 / 18	29 / 30	o_{20}
t_{26}	1	5 / 6	1	o_{21}
t_{28}	23 / 24	1	1	o_6
t_{29}	17 / 18	1	23 / 24	o_{12}
t_{31}	1	23 / 24	11 / 18	o_8
t_{32}	1	1	23 / 24	o_9
t_{34}	29 / 30	1	1	o_9
t_{35}	23 / 24	11 / 12	29 / 30	o_9
t_{37}	2 / 3	11 / 12	47 / 48	o_8
t_{38}	2 / 3	23 / 24	23 / 24	o_8
t_{40}	1	7 / 9	23 / 24	o_{19}
t_{41}	1	2 / 3	2 / 3	o_{23}

test data	$cpt\ (e_1, t_j)$	$cpt\ (e_2, t_j)$	$cpt\ (e_3, t_j)$	sol_{Kj}^{opt}
overall competence	25,33	24,92	24,47	

To keep an overview on the age of each entry, in the following resulting ***VKB*[4]** the time stamp has a background color with an increasing degree of grayness with increasing age (1 = 45%, 2 = 30 %, 3 = 15 %, 4 = 0%, i.e. white):

t_j	E_K	E_I	sol_{Kj}^{opt}	r_{ijk}	c_{ijk}	τ_s	D_C
t_1	e_1, e_3	$[e_1, e_2, e_3]$	o_6	[1,0,1]	[0,1,1]	1	
	e_3	$[e_1, e_2, e_3]$	o_4	[1,0,1]	[1,1,1]	3	
	e_2	$[e_1, e_2, e_3]$	o_{17}	[0,1,0]	[1,1,1]	4	
t_2	e_1, e_3	$[e_1, e_2, e_3]$	o_7	[0,0,1]	[0,0,1]	1	
	e_3	$[e_1, e_2, e_3]$	o_2	[1,0,1]	[1,1,1]	3	
	∅	$[e_1, e_2, e_3]$	o_2	[1,0,1]	[1,1,1]	4	
t_3	e_2	$[e_1, e_2, e_3]$	o_{20}	[0,1,0]	[0,1,1]	1	
	e_1	$[e_1, e_2, e_3]$	o_8	[1,0,1]	[1,1,1]	3	
t_4	e_1, e_3	$[e_1, e_2, e_3]$	o_2	[0,1,1]	[0,1,1]	1	
	e_1, e_3	$[e_1, e_2, e_3]$	o_2	[1,0,1]	[1,1,1]	3	
	e_1, e_3	$[e_1, e_2, e_3]$	o_2	[1,1,1]	[1,1,1]	4	
t_5	e_1, e_3	$[e_1, e_2, e_3]$	o_8	[1,1,1]	[1,0,1]	1	
	e_2	$[e_1, e_2, e_3]$	o_3	[1,1,1]	[1,1,1]	3	
	e_1	$[e_1, e_2, e_3]$	o_8	[1,1,1]	[1,1,1]	4	
t_6	e_1, e_2, e_3	$[e_1, e_2, e_3]$	o_9	[1,1,1]	[0,1,1]	1	
	e_1, e_2	$[e_1, e_2, e_3]$	o_9	[1,1,1]	[1,1,1]	3	
t_7	e_1, e_2, e_3	$[e_1, e_2, e_3]$	o_3	[1,1,1]	[0,1,1]	1	
	e_2	$[e_1, e_2, e_3]$	o_8	[1,1,1]	[1,1,1]	3	
	∅	$[e_1, e_2, e_3]$	o_2	[1,1,1]	[1,1,1]	4	
t_8	e_2, e_3	$[e_1, e_2, e_3]$	o_3	[1,1,1]	[0,1,1]	1	
	e_1, e_2	$[e_1, e_2, e_3]$	o_4	[1,1,1]	[1,1,1]	3	
	e_1, e_3	$[e_1, e_2, e_3]$	o_4	[1,1,1]	[1,1,1]	4	
t_9	e_1	$[e_1, e_2, e_3]$	o_1	[1,1,1]	[0,1,1]	1	
	e_1	$[e_1, e_2, e_3]$	o_6	[1,0,1]	[1,1,1]	3	
t_{10}	∅	$[e_1, e_2, e_3]$	o_1	[1,1,1]	[0,1,1]	1	
	e_1, e_3	$[e_1, e_2, e_3]$	o_2	[1,0,1]	[1,1,1]	3	
	e_1	$[e_1, e_2, e_3]$	o_2	[1,1,1]	[1,1,1]	4	
t_{11}	∅	$[e_1, e_2, e_3]$	o_1	[1,1,1]	[1,1,1]	1	
	e_2	$[e_1, e_2, e_3]$	o_8	[1,0,1]	[1,1,1]	3	
	e_2	$[e_1, e_2, e_3]$	o_3	[1,1,1]	[1,1,1]	4	
t_{12}	e_3	$[e_1, e_2, e_3]$	o_1	[1,1,1]	[1,1,1]	1	
	e_2	$[e_1, e_2, e_3]$	o_9	[1,1,1]	[1,1,1]	3	
t_{13}	e_2, e_3	$[e_1, e_2, e_3]$	o_9	[1,1,1]	[0,1,1]	1	
	∅	$[e_1, e_2, e_3]$	o_3	[1,1,1]	[1,1,1]	3	
	e_1	$[e_1, e_2, e_3]$	o_8	[1,1,1]	[1,1,1]	4	
t_{14}	e_3	$[e_1, e_2, e_3]$	o_3	[0,1,1]	[0,1,1]	1	
	e_1	$[e_1, e_2, e_3]$	o_9	[1,1,1]	[1,1,1]	3	

t_j	E_K	E_I	sol_{Kj}^{opt}	r_{ijk}	c_{ijk}	τ_s	D_C
	e_1 , e_2 , e_3	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,1]	4	
t_{15}	e_3	[e_1 , e_2 , e_3]	o_3	[0,1,1]	[0,1,1]	1	
	e_1 , e_3	[e_1 , e_2 , e_3]	o_3	[1,1,1]	[1,1,1]	2	
t_{16}	e_2	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,1]	1	
	e_1	[e_1 , e_2 , e_3]	o_3	[1,1,1]	[1,1,1]	2	
	e_2	[e_1 , e_2 , e_3]	o_8	[1,1,1]	[1,1,1]	4	
t_{17}	e_3	[e_1 , e_2 , e_3]	o_{20}	[1,1,1]	[0,1,1]	1	
	e_3	[e_1 , e_2 , e_3]	o_{20}	[1,1,1]	[1,1,0]	2	
	e_1 , e_2	[e_1 , e_2 , e_3]	o_{14}	[1,1,0]	[1,1,1]	4	
t_{18}	e_1 , e_3	[e_1 , e_2 , e_3]	o_6	[0,1,1]	[0,1,1]	1	
	e_3	[e_1 , e_2 , e_3]	o_6	[0,1,1]	[1,1,1]	2	
t_{19}	e_1 , e_3	[e_1 , e_2 , e_3]	o_{12}	[1,1,1]	[0,1,1]	1	
	e_3	[e_1 , e_2 , e_3]	o_{12}	[1,0,1]	[1,1,1]	2	
	e_1 , e_2	[e_1 , e_2 , e_3]	o_{17}	[1,1,1]	[1,1,0]	4	
t_{20}	e_2	[e_1 , e_2 , e_3]	o_{21}	[1,1,0]	[0,1,1]	1	
	e_3	[e_1 , e_2 , e_3]	o_4	[1,0,1]	[1,1,1]	2	
	e_1	[e_1 , e_2 , e_3]	o_{12}	[1,0,1]	[1,1,1]	4	
t_{21}	e_2	[e_1 , e_2 , e_3]	o_{21}	[1,1,1]	[1,1,1]	1	
	e_1	[e_1 , e_2 , e_3]	o_{15}	[1,1,1]	[1,1,1]	2	
t_{22}	e_1 , e_3	[e_1 , e_2 , e_3]	o_{17}	[1,1,1]	[1,1,1]	1	
	e_1	[e_1 , e_2 , e_3]	o_{17}	[1,1,1]	[1,1,1]	2	
	e_3	[e_1 , e_2 , e_3]	o_{20}	[1,1,1]	[1,1,1]	4	
t_{23}	e_3	[e_1 , e_2 , e_3]	o_{15}	[1,1,1]	[0,0,1]	1	
	e_3	[e_1 , e_2 , e_3]	o_{20}	[1,1,1]	[1,1,1]	2	
	∅	[e_1 , e_2 , e_3]	o_{20}	[1,1,1]	[1,1,1]	4	
t_{24}	∅	[e_1 , e_2 , e_3]	o_{15}	[1,1,1]	[0,1,1]	1	
	e_2	[e_1 , e_2 , e_3]	o_{16}	[1,1,1]	[1,1,1]	2	
t_{25}	e_1 , e_3	[e_1 , e_2 , e_3]	o_{20}	[1,1,1]	[1,1,1]	1	
	e_1 , e_3	[e_1 , e_2 , e_3]	o_{20}	[1,1,1]	[1,1,1]	2	
	e_1	[e_1 , e_2 , e_3]	o_{20}	[1,1,1]	[1,1,1]	4	
t_{26}	e_1 , e_2	[e_1 , e_2 , e_3]	o_{16}	[1,1,1]	[0,1,1]	1	
	e_1	[e_1 , e_2 , e_3]	o_{21}	[1,1,1]	[1,1,1]	2	
	e_3	[e_1 , e_2 , e_3]	o_{21}	[1,1,1]	[1,1,1]	4	
t_{27}	e_2	[e_1 , e_2 , e_3]	o_{16}	[1,1,1]	[0,1,1]	1	
	e_1 , e_3	[e_1 , e_2 , e_3]	o_{22}	[1,1,1]	[1,1,1]	2	
t_{28}	e_2	[e_1 , e_2 , e_3]	o_8	[1,1,1]	[0,1,1]	1	
	e_2	[e_1 , e_2 , e_3]	o_8	[1,1,1]	[1,1,1]	2	
	e_3	[e_1 , e_2 , e_3]	o_6	[1,1,1]	[1,1,1]	4	
t_{29}	e_2	[e_1 , e_2 , e_3]	o_8	[1,1,1]	[1,1,0]	2	
	e_2	[e_1 , e_2 , e_3]	o_8	[1,1,1]	[1,1,1]	3	
	e_3	[e_1 , e_2 , e_3]	o_{12}	[1,1,1]	[1,1,1]	4	
t_{30}	e_2	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,1]	2	
	e_1	[e_1 , e_2 , e_3]	o_4	[1,1,1]	[1,1,1]	3	
t_{31}	e_1 , e_2 , e_3	[e_1 , e_2 , e_3]	o_2	[0,1,1]	[1,1,1]	2	

t_j	E_K	E_I	sol_{Kj}^{opt}	r_{ijk}	c_{ijk}	τ_s	D_C
	e_1	[e_1 , e_2 , e_3]	o_8	[1,1,1]	[1,1,1]	3	
	e_1	[e_1 , e_2 , e_3]	o_8	[1,1,1]	[1,1,1]	4	
t_{32}	e_2	[e_1 , e_2 , e_3]	o_3	[1,1,1]	[1,1,1]	2	
	e_2 , e_3	[e_1 , e_2 , e_3]	o_3	[1,1,1]	[1,0,1]	3	
	e_1	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,1]	4	
t_{33}	e_2	[e_1 , e_2 , e_3]	o_8	[1,1,1]	[1,1,1]	2	
	e_2	[e_1 , e_2 , e_3]	o_8	[1,1,1]	[1,1,1]	3	
t_{34}	e_2	[e_1 , e_2 , e_3]	o_2	[1,1,1]	[1,1,1]	2	
	e_3	[e_1 , e_2 , e_3]	o_4	[1,1,1]	[1,1,1]	3	
	e_2	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,1]	4	
t_{35}	e_2	[e_1 , e_2 , e_3]	o_8	[1,1,1]	[1,1,1]	2	
	e_2	[e_1 , e_2 , e_3]	o_8	[1,1,1]	[1,1,1]	3	
	e_2	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,1]	4	
t_{36}	e_2	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,1]	2	
	e_2	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,0]	3	
t_{37}	e_2	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,1]	2	
	e_2	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,1]	3	
	e_2	[e_1 , e_2 , e_3]	o_8	[1,1,1]	[1,1,1]	4	
t_{38}	e_2	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,1]	2	
	e_2	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,1]	3	
	e_2	[e_1 , e_2 , e_3]	o_8	[1,1,1]	[1,1,1]	4	
t_{39}	e_2	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,1]	2	
	e_2	[e_1 , e_2 , e_3]	o_9	[1,1,1]	[1,1,1]	3	
t_{40}	e_1 , e_2 , e_3	[e_1 , e_2 , e_3]	o_{23}	[1,1,1]	[1,1,1]	2	
	e_1	[e_1 , e_2 , e_3]	o_{19}	[1,1,1]	[1,0,1]	3	
	e_1 , e_3	[e_1 , e_2 , e_3]	o_{19}	[1,1,1]	[1,0,1]	4	
t_{41}	e_1	[e_1 , e_2 , e_3]	o_{19}	[1,1,1]	[1,1,1]	2	
	∅	[e_1 , e_2 , e_3]	o_{19}	[1,1,1]	[1,1,1]	3	
	e_1	[e_1 , e_2 , e_3]	o_{23}	[1,1,1]	[1,0,1]	4	
t_{42}	e_1 , e_2 , e_3	[e_1 , e_2 , e_3]	o_{23}	[1,1,1]	[1,1,1]	2	
	e_1 , e_2 , e_3	[e_1 , e_2 , e_3]	o_{23}	[1,1,1]	[1,1,1]	3	

Again, this setting of the experiment there is never a need to compute a “most similar expert”, because each expert solved or rated the considered case in former sessions. To validate this approach nevertheless, we include the calculation of a most similar expert for $e_3$and compare his/her replies with the one of $VESA_3^3$ as well:

		# of identical solutions	# of identical ratings	Σ per session	Σ at all
e_1	1st session	11	64	75	335
	2nd session	9	69	78	
	3rd session	3	82	85	
	4th session	4	93	97	
e_2	1st session	5	59	54	274
	2nd session	4	66	70	

	3rd session	1	59	60	
	4th session	2	88	90	

Thus, e_1 is the most similar expert to e_3.

The responses of $VESA_3^3$, to the requests concerning the external knowledge EK_4 and compared with the responses of its human origin e_3 look as follows. Identical behavior (solutions or ratings) of **VESA** and human origin is marked red.

EK_4	solution of			EK_4	solution of			EK_4	solution of			EK_4	solution of		
	$VESA_3^3$	e_3	e_1		$VESA_3^3$	e_3	e_1		$VESA_3^3$	e_3	e_1		$VESA_3^3$	e_3	e_1
t_1	o_6	o_2	o_6	t_{11}	o_3	o_{11}	o_{13}	t_{22}	o_{17}	o_{20}	o_{17}	t_{32}	**o_4**	**o_4**	o_9
t_2	o_7	o_5	o_7	t_{13}	o_5	o_{10}	o_8	t_{23}	**o_{16}**	**o_{16}**	o_{13}	t_{34}	o_9	o_8	o_{10}
t_4	**o_2**	**o_2**	o_2	t_{14}	**o_9**	**o_9**	o_9	t_{25}	o_{20}	o_{22}	o_{20}	t_{35}	**o_5**	**o_5**	o_{11}
t_5	o_8	o_5	o_8	t_{16}	o_{11}	o_7	o_{11}	t_{26}	**o_{21}**	**o_{21}**	o_{16}	t_{37}	o_5	o_{10}	o_{25}
t_7	o_3	o_{10}	o_3	t_{17}	o_1	o_6	o_{14}	t_{28}	**o_6**	**o_6**	o_9	t_{38}	o_{11}	o_4	o_{25}
t_8	**o_4**	**o_4**	o_4	t_{19}	o_{12}	o_3	o_{17}	t_{29}	o_{16}	o_{12}	o_3	t_{40}	o_{23}	o_{19}	o_{19}
t_{10}	o_2	o_9	o_2	t_{20}	o_{12}	o_1	o_{12}	t_{31}	o_2	o_7	o_8	t_{41}	o_{18}	o_{14}	o_{23}

EK_4	solution	rating of			certainty of			EK_4	solution	rating of			certainty of		
		$VESA_3^3$	e_3	e_1	$VESA_3^3$	e_3	e_1			$VESA_3^3$	e_3	e_1	$VESA_3^3$	e_3	e_1
t_1	o_2	**0**	**0**	1	1	0	1	t_{23}	o_{13}	1	0	1	0	1	1
	o_4	**1**	**1**	1	1	1	1		o_{14}	0	1	1	1	0	1
	o_6	**1**	**1**	1	1	1	1		o_{15}	0	1	1	1	1	1
	o_{17}	**0**	**0**	0	1	1	1		o_{16}	**1**	**1**	1	1	1	1
	o_{18}	1	0	0	0	1	1		o_{18}	1	0	0	1	1	1
t_2	o_2	**1**	**1**	1	1	1	1		o_{20}	**1**	**1**	1	1	1	1
	o_5	0	1	1	1	0	1	t_{25}	o_{15}	**1**	**1**	1	1	1	1
	o_7	**1**	**1**	0	1	1	1		o_{17}	0	1	1	1	1	1
	o_{15}	**0**	**0**	0	1	1	1		o_{20}	**1**	**1**	1	1	1	1
t_4	o_2	**1**	**1**	1	1	1	1		o_{22}	**1**	**1**	1	1	1	1
	o_{10}	**1**	**1**	1	1	1	1	t_{26}	o_{16}	**1**	**1**	1	1	1	1
	o_{20}	**0**	**0**	0	1	1	1		o_{18}	1	0	0	0	1	1
t_5	o_3	**1**	**1**	1	1	1	1		o_{21}	**1**	**1**	1	1	1	1
	o_5	0	1	1	1	1	1	t_{28}	o_2	0	1	1	1	0	1
	o_8	**1**	**1**	1	1	1	1		o_6	**1**	**1**	1	1	1	1
	o_{18}	**0**	**0**	0	1	1	1		o_8	**1**	**1**	1	1	1	1
t_7	o_2	**1**	**1**	1	1	1	1		o_9	0	1	1	1	1	1
	o_3	**1**	**1**	1	1	1	1		o_{16}	1	0	1	0	1	0
	o_6	0	1	1	1	0	1	t_{29}	o_3	0	1	1	1	0	1
	o_8	**1**	**1**	1	1	0	1		o_8	**1**	**1**	1	0	1	1
	o_{10}	0	1	1	1	0	1		o_{12}	**1**	**1**	1	1	1	1
t_8	o_3	**1**	**1**	1	1	1	1		o_{16}	**0**	**0**	1	0	1	0
	o_4	**1**	**1**	1	1	1	1	t_{31}	o_2	**1**	**1**	1	1	0	1
	o_9	0	1	1	1	1	1		o_7	**1**	**1**	0	1	1	1
	o_{10}	**1**	**1**	1	1	1	1		o_8	0	1	1	1	1	1
t_{10}	o_1	**1**	**1**	1	0	0	1		o_9	0	1	1	1	1	1
	o_2	**1**	**1**	1	1	1	1		o_{16}	1	0	1	1	1	1

EK_4	solution	rating of			certainty of		
		$VESA_3^3$	e_3	e_1	$VESA_3^3$	e_3	e_1
	o_9	0	1	1	1	1	1
t_{11}	o_1	0	1	1	0	0	1
	o_3	**1**	**1**	1	1	1	1
	o_8	**1**	**1**	1	1	1	1
	o_{11}	0	1	1	1	1	1
	o_{13}	**0**	**0**	0	1	1	1
t_{13}	o_3	**1**	**1**	1	1	0	1
	o_8	**1**	**1**	1	1	1	1
	o_9	**1**	**1**	1	1	1	1
	o_{10}	0	1	1	1	1	1
	o_{12}	0	1	0	1	1	1
t_{14}	o_3	**1**	**1**	1	1	1	1
	o_9	**1**	**1**	1	1	1	1
t_{16}	o_3	**1**	**1**	1	1	1	1
	o_7	**1**	**1**	0	1	1	1
	o_8	0	1	1	1	1	1
	o_9	**1**	**1**	1	1	1	1
	o_{11}	**1**	**1**	1	1	1	1
t_{17}	o_1	**1**	**1**	0	1	1	1
	o_6	**1**	**1**	0	1	1	1
	o_{14}	**0**	**0**	1	1	1	1
	o_{20}	1	0	1	0	1	1
t_{19}	o_1	**1**	**1**	0	0	1	1
	o_3	**1**	**1**	0	1	1	1
	o_{12}	**1**	**1**	0	1	1	1
	o_{17}	**0**	**1**	1	1	0	1
t_{20}	o_1	**1**	**1**	1	0	1	1
	o_4	**1**	**1**	1	1	1	1
	o_{12}	**1**	**1**	1	1	1	1
	o_{16}	**0**	**0**	0	1	1	1
	o_{21}	**0**	**0**	0	1	1	1
t_{22}	o_{14}	**1**	**1**	1	1	0	1
	o_{17}	**1**	**1**	1	1	1	1
	o_{20}	**1**	**1**	1	1	1	1

EK_4	solution	rating of			certainty of		
		$VESA_3^3$	e_3	e_1	$VESA_3^3$	e_3	e_1
t_{32}	o_3	**1**	**1**	1	1	1	1
	o_4	**1**	**1**	1	1	1	1
	o_8	0	1	1	1	1	1
	o_9	0	1	1	1	1	1
	o_{16}	1	0	1	0	1	1
t_{34}	o_2	**1**	**1**	1	1	0	0
	o_4	0	1	1	1	1	1
	o_8	**1**	**1**	1	1	1	1
	o_9	**1**	**1**	1	1	1	1
	o_{10}	0	1	1	1	1	1
	o_{17}	**0**	**0**	1	1	1	1
t_{35}	o_5	**1**	**1**	1	1	1	1
	o_8	**1**	**1**	1	1	1	1
	o_9	0	1	1	1	1	1
	o_{11}	0	1	1	1	1	1
	o_{17}	**0**	**0**	1	1	1	0
t_{37}	o_8	0	1	1	1	1	1
	o_9	**1**	**1**	1	1	1	1
	o_{10}	**1**	**1**	1	1	1	1
	o_{24}	0	*norating*	0	0	0	1
	o_{25}	**0**	**0**	1	1	1	1
t_{38}	o_4	**1**	**1**	1	1	1	1
	o_8	0	1	1	1	1	1
	o_9	**1**	**1**	1	1	1	1
	o_{24}	0	*norating*	0	0	0	1
	o_{25}	**0**	**0**	1	1	1	1
t_{40}	o_{18}	0	1	1	1	1	1
	o_{19}	**1**	**1**	1	1	1	1
	o_{22}	**1**	**1**	0	1	1	1
	o_{23}	**1**	**1**	1	1	1	1
t_{41}	o_{14}	**1**	**1**	0	1	1	1
	o_{19}	**1**	**1**	1	1	1	1
	o_{23}	0	1	1	1	1	1

Evaluation of the 4th session:

- the number a_4 of cases from VKB^3, which were subject of the rating session and relate it to $|EK_4|$: $A_4 := a_4 / |EK_4|$
  - *There were **24** (!) cases, for which VKB^3 had a solution which was not in the process anyway, i.e. $a_4 = 24$. This solution became worse marks in the 2nd session.* $A_4 := 24 / 28 \approx 0{,}85$
- the number b_4 of cases from VKB^3, which provided the optimal (best rated) solution and relate it to $|EK_4|$: $B_4 := b_4 / |EK_4|$

- *Two of the 24 cases that have been submitted by VKB3 became the optimal solution,* i.e. $b_4 = 2$: $\boldsymbol{B_4 := 2 / 28 \approx 0{,}071}$

- the number $\boldsymbol{c_4}$ of cases from $\boldsymbol{VKB^4}$, for which a new solution has been introduced into ***VKB*** and relate it to $\boldsymbol{|EK_3|}$: $\boldsymbol{C_4 := c_4 / |EK_4|}$

  - *For $c_4 = 17$ of the 28 cases in $\boldsymbol{EK_4}$ a new solution has been introduced in $\boldsymbol{VKB^3}$ towards $\boldsymbol{VKB^4}$:* $\boldsymbol{C_4 := 17 / 28 \approx 0{,}61}$

- the number $\boldsymbol{d_4}$ of solutions and ratings, which are identical responses of $\boldsymbol{e_3}$ and $\boldsymbol{VESA_3}$ and relate it to the number of required solutions and ratings: $\boldsymbol{D_4 := d_4 /}$ ***number of expert responses altogether***

  - *For only 8 (out of 28) cases $\boldsymbol{VESA_3}$ provided the same solution as its human origin.*
  - *For 82 (out of 122) rating requests $\boldsymbol{VESA_3}$ provided the same rating as its human origin.*
  - *Thus, $d_4 = (8+82) = 90$:* $\boldsymbol{D_4 := 90 / 150 \approx 0{,}6}$

7.5 Final Results: The Benefit of VKB and VESA

Answers to the vacant questions are as follows:

1. **Does a *VKB* contribute to the validation sessions in an increasing degree with an increasing number of validation sessions ($A_4 > A_3 > A_2$) ?**

   With $A_4 \approx 0{,}85$, $A_3 \approx 0{,}071$, and $A_2 \approx 0{,}071$ the unequation is almost met.

   The contribution effect could not really be expected right after the first or second session. A ***VKB*** needs to gain a certain amount of "historical experience", before it can contribute to a new session in a sufficient way.
   Indeed, after the third session a remarkable number (24 out of 28 possible cases) of ***VKB*** have been introduced in the rating process.
   A 5th, 6th and further sessions would show this effect much more convincingly.

2. **Does the *VKB* contribute valid knowledge (best rated solutions) in an increasing degree with an increasing number of validation sessions ($B_4 > B_3 > B_2$) ?**

   With $B_4 \approx 0{,}071$, $B_3 = 0$, and $B_2 = 0$ the unequation is not really met.

   However, in the 4th session for two cases VKB contributed solutions, which have not been provided by the human experts, but won the "rating contest", i.e. received better marks than any solution of the present humans.
   We tend to consider this as a success as well. Even just one such case is a noteworthy contribution towards a more and dependable system.

In fact, the more entries a VKB gains, the higher is the number of solutions which are subject of the rating process and therefore, the higher is the probability that such a solution enjoys to be the optimal one.

3. **Does the *VKB* skim the human expertise in an increasing degree with an increasing number of validation sessions ($C_4 < C_3 < C_2$) ?**

   With $C_4 \approx 0{,}61$, $C_3 \approx 0{,}57$, and $C_2 = 0{,}5$ the unequation is not met.

   Maybe we asked the wrong question. The underlying assumption for this question is a static problem domain with a static domain knowledge, which needs to be explored systematically.
   We should have known, that this is true for most interesting problem domains. Both is subject of change over time the domain knowledge itself and its reflection in human brains.

4. **Do the *VESA*s really model their human origin in an increasing degree with an increasing number of validation sessions ($D_4 > D_3 > D_2$) ?**

   With $D_4 = 0{,}6$, $D_3 \approx 0{,}63$, and $D_2 = 0{,}51$ we can at least claim that $D_4 \geq D_3 \geq D_2$ is almost met.

   However, in the setting of the experiment a ***VESA*** was always based on **former considerations** of a present case **by the same expert**. A view on the decisions of the "most similar expert" (which is provided as a "gray column" in the associated tables) shows, that this situation was better, if we have had a setting where a former solution or rating is not available.

   The fact that these numbers are not convincing is based on a human factor in the experiment and the approach itself:
   - All experts changed their opinion during the experiments for a remarkable number of cases. We believe, the basic reasons for this are
     - the interpretation of the cases itself and
     - the fact that a solution not exclusively depends on the provided input attributes.
   - In particular the rating process of a ***VESA*** on the bases of a last consideration of this case in a solving (not rating) session is based on the assumption the domain is deterministic by nature. This is certainly not true for most interesting problem domains, in particular in AI application fields. An expert provides "a preferable" solution if he/she is asked to provide a one, but rates other solutions as (also) correct in the rating process. An example along with a refinement suggestion for ***VESA*** is given in the next section.

7.6 Derived Improvements of VKB and VESA

7.6.1 Improvements of VKB

The VKB already includes all aspects of "collective historical experience" that has been provided by former expert panels. There is no need to collect more or other validation knowledge in the ***VKB*** to improve its benefit.

The only issue, that might be a subject of further refinement is the issue of "outdating" knowledge.
Since the number of solutions that are likely to be introduced in the rating process increases with the number of sessions, the human workload to rate all these solutions might become a costly factor.
A concept to weight former optimal solutions with a "rating necessity degree" depending on (1) how often this solution occurred in the past and (2) how long this is ago is certainly a way to face this problem.

7.6.2 Improvements of VESA

The VESA approach, on the other hand, needs a general revision. The issues that needs to be considered towards a "***next generation VESA***" are as follows:

Considerations of non-deterministic problem domains by VESA

If ***VESA*** is requested to provide a solution and the latest consideration of its human origin was a rating, it is quite arbitrary, which of the alternative solutions has been rated by him/her. The human origin might rate a (last considered) solution as (also) "correct", but provide in case of being requested to solve the case, a different solution.
Example: The first author would always prefer an Italian dry white wine (*Orivieto* or *Soave, e.g.*) along with some (only) vegetable dish and provide this as solution to this case. If he is requested to rate the solution *Chardonnay* , I would rate it with *r:=1* and *c:= 1* neverthless, because this is also a quite good choice.
Thus, modeling a solution behavior by considering former ratings might not be the right way.

Vice versa, if ***VESA*** is requested to provide a rating based on a latest consideration of its human origin and this consideration is a test case solution, ***VESA*** assumes that the solution formerly provided is the **only one** correct solution.
This reflected by providing a rating *r:=1* for this solution and *r:=0* for any other alternative solution.

Both does not reflect the reality of most interesting problem domains, especially in AI application domains.

To overcome this drawback, ***VESA*** needs to be refined in the following manner:

- If ***VESA*** is requested to provide a solution, only the former solution behavior of its human origin (not the rating behavior) should be considered. In case there is no former solution, the approach with the "most similar" human expert needs to be applied.
- Vice versa, if ***VESA*** is requested to provide a rating, only the former rating behavior of its human origin (not the solution behavior) should be considered. In case there is no former rating, the approach with the "most similar" human expert needs to be applied.

Computation of a most similar expert

The similarity of opinions based on both identical solutions and identical ratings ratings led to the same most similar expert as computing the similarity only based on identical solutions, but counting only the identical solutions makes differences more

explicit. It might be reasonable, to define the degree of similarity based on only the number of identical solutions provided to the same test data in the same session.

It turned out to be likely, that the computation of a most similar expert ends up with several experts with the same degree of similarity with respect to their former responses.
In this case, we suggest to prefer the latest sessions. This seems to be reasonable, because also these similarities are subject of natural change. In fact, these changes are reflected by the absolute amount of identical responses as well former approach as well sooner or later, but in case of the situation considered here this is reasonable.

Nevertheless, the most similar expert provided too often results different from the human original that was modeled by it. The use of former responses of the human original revealed much more identical replies with the current responses.
In the setting of this experiment, there was no need to use this approach, because there was a former reply of the real human origin. However, the concept of similarity with other humans might be a subject of revision in future research.

Permanent validation of VESA

As discussed in section 4, there was some serious doubt in the usefulness of ***VESA***. As a result, the authors agreed to analyze the experimentation results to validate ***VESA***s "validation knowledge".

In fact, this validation needs to be performed in the regular use of ***VESA*** as well by employing a ***VESA*** all the time, even if its human origin is available. By comparing ***VESA***s responses with the one of its human origin a statement of ***VESA***s quality can be derived.
If we can quantify this quality, a minimum value needs to be defined to ensure a certain quality of ***VESA***s knowledge. If the value can not be guarantied, the use of ***VESA*** as a substitute of its human origin needs to be rejected.

Completion of VESA towards other than (former) test cases

The fact, that a ***VESA*** can only provide validation knowledge (solutions, ratings) to cases that have been test cases in former (solving or rating) sessions, turned out to be a limitation that questions the practical value of the concept.
Test cases of an actual session are often been different from test cases that have been considered in former sessions.
Following the intension of modeling the individual human expertise of its human origin, the ***VESA*** approach needs to be refined by a concept of a "most likely" response of this human origin in case there is no "most similar" expert who ever considered an actual case in the past.

8 Summary and Conclusion

Generally, the idea of ***VKB*** is certainly the appropriate way to establish new sources of knowledge for system validation towards more dependable systems.

On a very first view, ***VKB*** seems to enlarge the number of solutions significantly with an creasing age (number of sessions that served to maintain VKB), but this is partly due to the experimental conditions: Since we omitted the System Refinement step of

the technology (because of time limitation and non-significance for the purpose of the experiment), the system solution was never adapted to the insights of each validation cycle. Thus, a solution which turned out to be invalid was never replaced in the rule base and therefore, always a subject of validation in each cycle. Since this would not happen in practice, we can assume that in many test cases the number of solutions that need to be rated can be decreased by one.
However, this problem needs to be faced before using a VKB in a commercial application contexts. Appropriate ideas are derived and sketched in this report.

Whether or not the ***VESA*** approach might become a subject of technology transfer into a real world application context, can not be derived from the experiment results.
At least the experiment revealed some weaknesses of the approach and the issues which needs to be addressed in the development of "next generation ***VESA***s":

- considerations of non-deterministic problem domains by ***VESA***,

the computation of a most similar expert, which really reflects similarities in "thinking structures" of humans,
a concept to validate ***VESA*** permanently (even if its human origin is available),
a concept to employ ***VESA***s also for other cases than former test cases.

In fact, the experiment itself was a valuable source of knowledge. We gained a lot of insights about the effects of our conceptual ideas and developed first refinement and revision ideas towards AI systems with a better performance.
The authors are convinced, that the general approach of permanently checking the systems against cases derived from (historical and present) practice, is a necessary contribution to face the current problems of system dependability.

In an information environment, that allows (and sometimes even forces) us to build more and more complex systems that we are not able to completely control
with respect to the development process itself, because there is no really complete model of the application requests and the functionality of the development tools and
with respect to the use of the system within an application environment,
the development, test and refinement of concepts for a permanent system validation becomes a key issue for further progress.

9 References

(Knauf 2000)
Knauf, Rainer: *Validating Rule based Systems – A Complete Methodology*. Aachen, Germany: Shaker, 2000 (Berichte aus der Informatik) Zugl.: Ilmenau, Technische Universität, Habilitationsschrift, ISBN 3-8265-8293-4, 2000.

(Knauf et al. 2002)
Knauf, Rainer; Gonzalez, Avelino J.; Abel, Thomas: A Framework for Validation of Rule-Based Systems. In: *IEEE Transactions of Systems, Man and Cybernetics* - Part B: Cybernetics, Volume 32, No. 3, June 2002, pp. 281-295, 2002.

(Knauf et al. 2003)
Knauf, Rainer; Gonzalez, Avelino J.; Tsuruta, Setsuo: Utilizing Validation Experience for System Validation. Russell and Haller (eds.): *Proceedings of 16th International Florida Artificial Intelligence Research Society Conference*

2003 (FLAIRS-2003), St. Augustine, FL, USA, pp. 223-227, Menlo Park, CA: AAAI Press, 2003.

(Knauf et al. 2004a)

Knauf, Rainer; Tsuruta, Setsuo; Ihara, Hirokazu; Gonzalez, Avelino J.; Kurbad, Torsten: Improving AI Systems' Dependability by Utilizing Historical Knowledge. accepted paper. *Proceedings of 10th International Symposium Pacific Rim Dependable Computing* (PRDC-2004), Papeete, Tahiti, French Polynesia, March 3-5, IEEE Computer Society Press, 2004.

(Knauf et al. 2004b)

Knauf, Rainer; Tsuruta, Setsuo; Uehara, Kenichi; Onoyama, Takashi; Kurbad, Torsten: The Power of Experience: On the Usefulness of Validation Knowledge.

accepted paper. V. Barr & Z. Markov (eds.): *Proceedings of 17th International Florida Artificial Intelligence Research Society Conference 2004* (FLAIRS-2004),

Miami Beach, FL, USA, Menlo Park, CA, USA: AAAI Press, 2004.

(Kurbad 2003)

Kurbad, Torsten: *A Concept to Apply a Turing Test Technology for System Validation that Utilizes External Validation Knowledge.* Diploma Thesis, Ilmenau Technical University, Germany, Faculty of Computer Science and Automation, # 2003-09-03/053/IN95/2238, 2003.

(Tsuruta et al. 2000a)

Tsuruta, Setsuo; Onoyama, Takashi; Kubota, S; Oyanagi, K. : Validation Method for Intelligent Systems. Etheredge/Manaris (eds): *Proceedings of 13th International Florida Artificial Intelligence Research Society Conference 2000* (FLAIRS-2000), pp. 361-365, Menlo Park, CA, USA: AAAI Press, 2004.

(Tsuruta et al. 2000b)

Tsuruta, Setsuo; Onoyama, Takashi; Kubota, S; Oyanagi, K. : Knowledge-based Approach for Validating Intelligent Systems. Kern (ed.): *Proceedings of 45th International Scientific Colloquium* (IWK-2000), Ilmenau, Germany, Technical University of Ilmenau, pp. 769-774, 2000.

(Tsuruta et al. 2002)

Tsuruta, Setsuo; Onoyama, Takashi; Taniguchi, Y.: Knowledge-Based Validation Method for Validating Intelligent Systems. Kohlen (ed.): *Proceedings of the 15th International Florida Artificial Intelligence Research Society Conference 2002* (FLAIRS-2002), Pensacola Beach, FL, USA, pp. 226-230, Menlo Park, CA, USA: AAAI Press, 2002.

(Turing 1950)

Turing, A: Computing Machinery and Intelligence. Mind, LIX, # 236, pp. 433-460, 1950.

(Weiss 1997)

Weiss, G. (ed.): *Distributed Artificial Intelligence Meets Machine Learning: Learning in Multi-Agent Environments*. Berlin: Springer-Verlag, Lecture Notes in Computer Science Volume 1237, ISBN 3-540-62934-3.

(Singh and Huhns 1997)

Singh, M.; Huhns, M.: Challenges for Machine Learning in Cooperative Information Systems. In (Weiss 1997), pp. 11-24.

10 Appendix A

10.1 Questionnaire of the First Test Case Solution Session

in the field of suggesting an appropriate wine to a given meal

You are asked to suggest an appropriate wine to a given meal.

- If your suggestion is an item of the following list, please, fill in the number in the questionnaire table.
- In case you suggest a different one, please, describe it verbally.

List of possible suggestions

1. *Red wine, fruity, low tannin, less compound*
2. *Red wine, young, rich of tannin*
3. *Red wine, dark, fruity, from the „new world“*
4. *Red wine, maturely, from the Rhone valley (France)*
5. *Red wine, velvet, low tannin*
6. *Pinot Noire*
7. *Amarone*
8. *Burgundy, mature*
9. *Bordeaux, mature*
10. *Barolo, mature*
11. *Brunello, mature*
12. *Beaujolais*
13. *Rosé, dry, fruity, low tannin*
14. *White wine, light, fresh, low acid*
15. *White wine, strong, low tannin*
16. *White wine, rich in content*
17. *White wine, dry, fruity*
18. *Muscatel*
19. *Gewürztraminer*
20. *Sauvignon Blanc*
21. *Riesling, semi dry*
22. *Riesling, rich of acid*
23. *Port wine*
24. *Any wine besides smooth one*

Again, feel free to suggest another kind of wine that listed up above.

To compute more general rules, the meals are described rather general, i.e. not as on a menu card of a restaurant.

Meal #	main ingredient	kind of preparation	style of preparation	suggested wine *(# of the above list or other nomination)*
1	pork	boiled	Asian	
2	pork	grilled	any	
3	pork	fried	any	
4	pork	stewed	any	
5	beef	boiled	Asian	
6	beef	grilled	any	
7	beef	fried	any	
8	beef	stewed	any	
9	veal	boiled	any	
10	veal	grilled	any	
11	veal	fried	any	
12	veal	stewed	any	
13	venison	boiled	any	
14	venison	grilled	any	
15	venison	fried	any	
16	venison	stewed	any	
17	fowl	boiled	any	
18	fowl	grilled	any	
19	fowl	fried	any	
20	fowl	stewed	any	
21	fish	non (raw)	Asian	
22	fish	steamed	Western	
23	fish	boiled	Asian	
24	fish	grilled	any	
25	fish	fried	any	
26	fish	stewed	Asian	
27	fish	deep fried	Asian	
28	hard cheese	non (raw)	Western	

10.2 Questionnaire of the first test case rating session
in the field of suggesting an appropriate wine to a given meal

Dear participant,

thank you very much for your kind suggestions for appropriate wines to given meals. In this session, we'd like know your rating of other peoples' suggestions to the same cases.
In the column "***correct?***" you can express your agreement to the considered suggestion by filling in "yes" or "no".
In the column "***sure?***" you can express some latent internal doubt in this rating by filling in "yes" or "no".

Meal #	main ingredient	kind of preparation	style of preparation	suggested wine	correct?	sure?
1	pork	boiled	Asian	Muscatel		
				Pinot Noir		
				Gewürztraminer		
2	pork	grilled	any	red wine, young, rich of tannin		
				Amarone		
				Riesling, semi dry		
				white wine, strong, low tannin		
3	pork	fried	any	red wine, young, rich of tannin		
				Burgundy, mature		
				Sauvignon Blanc		
				white wine, dry, fruity		
4	pork	stewed	any	Barolo, mature		
				red wine, young, rich of tannin		
				red wine, fruity, low tannin, less compound		
				Sauvignon Blanc		
5	beef	boiled	Asian	Muscatel		
				Burgundy, mature		
				red wine, dark, fruity, from the "new world"		
				red wine, fruity, low tannin, less compound		

Meal #	main ingredient	kind of preparation	style of preparation	suggested wine	correct?	sure?
6	beef	grilled	any	red wine, young, rich of tannin		
				Bordeaux, mature		
				red wine, dark, fruity, from the “new world”		
7	beef	fried	any	red wine, young, rich of tannin		
				red wine, dark, fruity, from the “new world”		
8	beef	stewed	any	Barolo, mature		
				red wine, mature, from the Rhone valley		
				red wine, dark, fruity, from the “new world”		
				red wine, velvet, low tannin		
9	veal	boiled	any	red wine, fruity, low tannin, less compound		
				Sauvignon Blanc		
				Rosé, dry, fruity, low tannin		
10	veal	grilled	any	red wine, fruity, low tannin, less compound		
				red wine, young, rich of tannin		
				Beaujolais		
11	veal	fried	any	red wine, fruity, low tannin, less compound		
				Rosé, dry, fruity, low tannin		
				red wine, mature, from the Rhone valley		
				Beaujolais		

Meal #	main ingredient	kind of preparation	style of preparation	suggested wine	correct?	sure?
12	veal	stewed	any	red wine, fruity, low tannin, less compound		
				red wine, velvet, low tannin		
				red wine, mature, from the Rhone valley		
13	venison	boiled	any	red wine, dark, fruity, from the “new world”		
				Burgundy, mature		
				Bordeaux, mature		
				red wine, young, rich of tannin		
14	venison	grilled	any	red wine, dark, fruity, from the “new world”		
				Bordeaux, mature		
15	venison	fried	any	red wine, dark, fruity, from the “new world”		
				Barolo, mature		
				Bordeaux, mature		
16	venison	stewed	any	Barolo, mature		
				Brunello, mature		
				Bordeaux, mature		
				red wine, young, rich of tannin		
17	fowl	boiled	any	red wine, fruity, low tannin, less compound		
				Muscatel		
				Sauvignon Blanc		

Meal #	main ingredient	kind of preparation	style of preparation	suggested wine	correct?	sure?
18	fowl	grilled	any	red wine, fruity, low tannin, less compound		
				Pinot Noir		
				Sauvignon Blanc		
				Rosé, dry, fruity, low tannin		
19	fowl	fried	any	red wine, fruity, low tannin, less compound		
				Beaujolais		
				Riesling, semi dry		
				Rosé, dry, fruity, low tannin		
20	fowl	stewed	any	red wine, fruity, low tannin, less compound		
				Riesling, semi dry		
				Sauvignon Blanc		
21	fish	non (raw)	Asian	Muscatel		
				white wine, rich in content		
				Riesling, semi dry		
				white wine, light, fresh, low acid		
22	fish	steamed	Western	white wine, light, fresh, low acid		
				white wine, dry, fruity		
23	fish	boiled	Asian	Muscatel		
				Rosé, dry, fruity, low tannin		
				white wine, rich in content		
				white wine, strong, low tannin		

Meal #	main ingredient	kind of preparation	style of preparation	suggested wine	correct?	sure?
24	fish	grilled	any	white wine, strong, low tannin		
				Gewürztraminer		
				white wine, rich in content		
				white wine, dry, fruity		
25	fish	fried	any	white wine, strong, low tannin		
				Sauvignon Blanc		
				white wine, rich in content		
				Rosé, dry, fruity, low tannin		
26	fish	stewed	Asian	Muscatel		
				white wine, rich in content		
				white wine, light, fresh, low acid		
27	fish	deep fried	Asian	Muscatel		
				white wine, rich in content		
				Rosé, dry, fruity, low tannin		
28	hard cheese	non (raw)	Western	white wine, rich in content		
				Bordeaux, mature		
				Burgundy, mature		
				red wine, dark, fruity, from the “new world”		

10.3 Questionnaire of the 2nd test case solution session
in the field of suggesting an appropriate wine to a given meal

Dear participant,

again, you are asked to suggest an appropriate wine to a given meal.

- If your suggestion is an item of the following list, please, fill in the number in the questionnaire table.
- In case you suggest a different one, please, describe it verbally.

List of possible suggestions

1. *Red wine, fruity, low tannin, less compound*
2. *Red wine, young, rich of tannin*
3. *Red wine, dark, fruity, from the „new world“*
4. *Red wine, maturely, from the Rhone valley (France)*
5. *Red wine, velvet, low tannin*
6. *Pinot Noire*
7. *Amarone*
8. *Burgundy, mature*
9. *Bordeaux, mature*
10. *Barolo, mature*
11. *Brunello, mature*
12. *Beaujolais*
13. *Rosé, dry, fruity, low tannin*
14. *White wine, light, fresh, low acid*
15. *White wine, strong, low tannin*
16. *White wine, rich in content*
17. *White wine, dry, fruity*
18. *Muscatel*
19. *Gewürztraminer*
20. *Sauvignon Blanc*
21. *Riesling, semi dry*
22. *Riesling, rich of acid*
23. *Port wine*
24. *Any wine besides smooth one*

Again, we encourage you to suggest another kind of wine than listed up above! New knowledge is very welcome and essential for the desired experimentation result!

To compute more general rules, the meals are described rather general, i.e. not as on a menu card of a restaurant.

Meal #	main ingredient	kind of preparation	style of preparation	suggested wine *# of the above list **or other nomination** (latter is very welcome!)*
15	venison	fried	any	
16	venison	stewed	any	
17	fowl	boiled	any	
18	fowl	grilled	any	
19	fowl	fried	any	
20	fowl	stewed	any	
21	fish	non (raw)	Asian	
22	fish	steamed	Western	
23	fish	boiled	Asian	
24	fish	grilled	any	
25	fish	fried	any	
26	fish	stewed	Asian	
27	fish	deep fried	Asian	
28	hard cheese	non (raw)	Western	
29	hard cheese	casserole	Western	
30	hard cheese	deep fried	Western	
31	soft cheese	non (raw)	Western	
32	soft cheese	casserole	Western	
33	soft cheese	deep fried	Western	
34	goat cheese	non (raw)	Western	
35	goat cheese	casserole	Western	
36	goat cheese	deep fried	Western	
37	blue mold cheese	non (raw)	Western	
38	blue mold cheese	casserole	Western	
39	blue mold cheese	deep fried	Western	
40	fruit dessert	non (raw)	any	
41	aromatic dessert	non (raw)	any	
42	ice cream	non (raw)	any	

10.4 Questionnaire of the 2nd test case rating session
in the field of suggesting an appropriate wine to a given meal

Dear participant,

thank you very much for your kind suggestions for appropriate wines to given meals. In this session, we'd like know your rating of other peoples' suggestions to the same cases.

In the column "***correct?***" you can express your agreement to the considered suggestion by filling in "yes" or "no". In the column "***sure?***" you can express some latent internal doubt in this rating by filling in "yes" or "no".

Additionally, you have the opportunity to express, that you can not provide any rating to this solution by filling in "no rating".

Meal #	main ingredient	kind of preparation	style of preparation	suggested wine	correct?	sure?
15	venison	fried	any	red wine, dark, fruity, from the "new world"		
				Bordeaux, mature		
16	venison	stewed	any	red wine, dark, fruity, from the "new world"		
				Bordeaux, mature		
				Brunello, mature		
17	fowl	boiled	any	red wine, fruity, low tannin, less compound		
				white wine, light, fresh, low acid		
				Muscatel		
				Sauvignon Blanc		
18	fowl	grilled	any	red wine, fruity, low tannin, less compound		
				Pinot Noire		
				Muscatel		
				Sauvignon Blanc		
19	fowl	fried	any	red wine, fruity, low tannin, less compound		
				Beaujolais		

				Rosé, dry, fruity, low tannin		
				Riesling, semi dry		
20	fowl	stewed	any	red wine, fruity, low tannin, less compound		
				red wine, mature, from the Rhone valley		
				Beaujolais		
				Riesling, semi dry		
21	fish	non (raw)	Asian	white wine, strong, low tannin		
				Muscatel		
				Riesling, semi dry		
				Riesling, rich of acid		
22	fish	steamed	Western	white wine, light, fresh, low acid		
				white wine, dry, fruity		
				Muscatel		
23	fish	boiled	Asian	Rosé, dry, fruity, low tannin		
				white wine, strong low tannin		
				white wine, rich in content		
				Muscatel		
				Sauvignon Blanc		
24	fish	grilled	any	white wine, strong, low tannin		
				white wine, rich in content		
				Gewürztraminer		
25	fish	fried	any	white wine, strong, low tannin		
				white wine, rich in content		
				Sauvignon Blanc		
26	fish	stewed	Asian	white wine, rich in content		
				Muscatel		

				Sauvignon Blanc		
				Riesling, semi dry		
27	fish	deep fried	Asian	white wine, rich in content		
				Muscatel		
				Riesling, rich of acid		
28	hard cheese	non (raw)	Western	Pinot Noire		
				Burgundy, mature		
				white wine, rich in content		
29	hard cheese	casserole	Western	Burgundy, mature		
				white wine, strong, low tannin		
				white wine, rich in content		
30	hard cheese	deep fried	Western	Bordeaux, mature		
				white wine, rich in content		
31	soft cheese	non (raw)	Western	red wine, young, rich of tannin		
				white wine, rich in content		
32	soft cheese	casserole	Western	red wine, young, rich of tannin		
				red wine, dark, fruity, from the “new world”		
				red wine, mature, from the Rhone valley		
				white wine, rich in content		
33	soft cheese	deep fried	Western	red wine, young, rich of tannin		
				red wine, dark, fruity, from the “new world”		
				Burgundy, mature		
				white wine, rich in content		
34	goat cheese	non (raw)	Western	red wine, young, rich of tannin		
				red wine, velvet, low tannin		
				Bordeaux, mature		

				white wine, dry, fruity		
35	goat cheese	casserole	Western	red wine, velvet, low tannin		
				Burgundy, mature		
				Barolo, mature		
				white wine, dry, fruity		
36	goat cheese	deep fried	Western	red wine, velvet, low tannin		
				Bordeaux, mature		
				Brunello, mature		
				white wine, dry, fruity		
37	blue mold cheese	non (raw)	Western	red wine, young, rich of tannin		
				Bordeaux, mature		
				Barolo, mature		
				any wine besides smooth one		
38	blue mold cheese	casserole	Western	red wine, young, rich of tannin		
				Bordeaux, mature		
				Beaujolais		
				any wine besides smooth one		
39	blue mold cheese	deep fried	Western	Bordeaux, mature		
				any wine besides smooth one		
40	fruit dessert	non (raw)	any	Riesling, rich of acid		
				Port wine		
41	aromatic dessert	non (raw)	any	Muscatel		
				Gewürztraminer		
				Riesling, rich of acid		
42	ice cream	non (raw)	any	Port wine		

10.5 Questionnaire of the 3rd test case solution session
in the field of suggesting an appropriate wine to a given meal

Dear participant,

again, you are asked to suggest an appropriate wine to a given meal.

- If your suggestion is an item of the following list, please, fill in the number in the questionnaire table.
- In case you suggest a different one, please, describe it verbally.

Additionally, you have the opportunity to express that you really can not suggest a wine, because of not having an idea by filling in "no solution".

List of possible suggestions

1. *Red wine, fruity, low tannin, less compound*
2. *Red wine, young, rich of tannin*
3. *Red wine, dark, fruity, from the „new world"*
4. *Red wine, maturely, from the Rhone valley (France)*
5. *Red wine, velvet, low tannin*
6. *Pinot Noire*
7. *Amarone*
8. *Burgundy, mature*
9. *Bordeaux, mature*
10. *Barolo, mature*
11. *Brunello, mature*
12. *Beaujolais*
13. *Rosé, dry, fruity, low tannin*
14. *White wine, light, fresh, low acid*
15. *White wine, strong, low tannin*
16. *White wine, rich in content*
17. *White wine, dry, fruity*
18. *Muscatel*
19. *Gewürztraminer*
20. *Sauvignon Blanc*
21. *Riesling, semi dry*
22. *Riesling, rich of acid*
23. *Port wine*
24. *Any wine besides smooth one*

Again, we encourage you to suggest another kind of wine than listed up above! New knowledge is very welcome and essential for the desired experimentation result!

To compute more general rules, the meals are described rather general, i.e. not as on a menu card of a restaurant.

Meal #	main ingredient	kind of preparation	style of preparation	suggested wine *# of the above list **or other nomination** (latter is very welcome!)*
1	pork	boiled	Asian	
2	pork	grilled	any	
3	pork	fried	any	
4	pork	stewed	any	
5	beef	boiled	Asian	
6	beef	grilled	any	
7	beef	fried	any	
8	beef	stewed	any	
9	veal	boiled	any	
10	veal	grilled	any	
11	veal	fried	any	
12	veal	stewed	any	
13	venison	boiled	any	
14	venison	grilled	any	
29	hard cheese	casserole	Western	
30	hard cheese	deep fried	Western	
31	soft cheese	non (raw)	Western	
32	soft cheese	casserole	Western	
33	soft cheese	deep fried	Western	
34	goat cheese	non (raw)	Western	
35	goat cheese	casserole	Western	
36	goat cheese	deep fried	Western	
37	blue mold cheese	non (raw)	Western	
38	blue mold cheese	casserole	Western	
39	blue mold cheese	deep fried	Western	
40	fruit dessert	non (raw)	any	
41	aromatic dessert	non (raw)	any	
42	ice cream	non (raw)	any	

10.6 Questionnaire of the 3rd test case rating session
in the field of suggesting an appropriate wine to a given meal

Dear participant,

thank you very much for your kind suggestions for appropriate wines to given meals. In this session, we'd like know your rating of other peoples' suggestions to the same cases.

In the column "***correct?***" you can express your agreement to the considered suggestion by filling in "*yes*" or "*no*". In the column "***sure?***" you can express some latent internal doubt in this rating by filling in "*yes*" or "*no*".
Additionally, you have the opportunity to express, that you can not provide any rating to this solution by filling in "*no rating*".

Since some meal descriptions raised some questions, the following specification of some terms might be helpful to interpret it:
- ***Fish*** means "white fish", not red one.
- ***Cheese*** means the one with the strongest taste in its category, because cheese with a light taste would not be a "main ingredient".
- ***Asian style*** means the more spicy variant of it. This is the more Southeast Asian style, less the Japanese one.
- ***Western style*** means just pepper and salt. Meat in this style goes along with brown sauce (not white one) and in case of fish it is supplemented by garlic.
- ***Any style*** means no spice besides salt and pepper. In case stewed food it also means a dark colored sauce.

Meal #	main ingredient	kind of preparation	style of preparation	suggested wine	correct?	sure?
1	pork	boiled	Asian	red wine, mature, from the Rhone Valley		
				Pinot Noire		
				Muscatel		
				Riesling, semi dry		
2	pork	grilled	any	red wine, young, rich of tannin		
				Amarone		
				Sauvignon Blanc		
3	pork	fried	any	red wine, young, rich of tannin		
				red wine, dark, fruity, from the "new world"		
				Burgundy, mature		
				Port wine		

4	pork	stewed	any	red wine, young, rich of tannin		
				Barolo, mature		
				any wine besides smooth one		
5	beef	boiled	Asian	red wine, fruity, low tannin, less compound		
				red wine, dark, fruity, from the “new world”		
				Burgundy, mature		
				Muscatel		
6	beef	grilled	any	red wine, young, rich of tannin		
				Bordeaux, mature		
7	beef	fried	any	red wine, young, rich of tannin		
				red wine, dark, fruity, from the “new world”		
				Burgundy, mature		
8	beef	stewed	any	red wine, young, rich of tannin		
				red wine, mature, from the Rhone Valley		
				Barolo, mature		
9	veal	boiled	any	red wine, fruity, low tannin, less compound		
				red wine, velvet, low tannin		
				Pinot Noire		
				white wine, strong, low tannin		
10	veal	grilled	any	red wine, fruity, low tannin, less compound		
				red wine, young, rich of tannin		
				red wine, dark, fruity, from the “new world”		
11	veal	fried	any	red wine, fruity, low tannin, less compound		

				red wine, dark, fruity, from the “new world”		
				Burgundy, mature		
				Rosé, dry, fruity, low tannin		
12	veal	stewed	any	red wine, fruity, low tannin, less compound		
				red wine, mature, from the Rhone Valley		
				red wine, velvet, low tannin		
				Bordeaux, mature		
13	venison	boiled	any	red wine, dark, fruity, from the “new world”		
				red wine, velvet, low tannin		
				Burgundy, mature		
				Bordeaux, mature		
14	venison	grilled	any	red wine, young, rich of tannin		
				red wine, dark, fruity, from the “new world”		
				red wine, mature, from the Rhone Valley		
				Bordeaux, mature		
29	hard cheese	casserole	Western	red wine, dark, fruity, from the “new world”		
				red wine, mature, from the Rhone Valley		
				Burgundy, mature		
				white wine, rich in content		
30	hard cheese	deep fried	Western	red wine, young, rich of tannin		
				red wine, mature, from the Rhone Valley		
				Bordeaux, mature		
				white wine, rich in content		

31	soft cheese	non (raw)	Western	red wine, young, rich of tannin		
				red wine, mature, from the Rhone Valley		
				Burgundy, mature		
				white wine, rich in content		
32	soft cheese	casserole	Western	red wine, dark, fruity, from the “new world”		
				Bordeaux, mature		
				white wine, rich in content		
33	soft cheese	deep fried	Western	red wine, dark, fruity, from the “new world”		
				Burgundy, mature		
				Bordeaux, mature		
				white wine, rich in content		
34	goat cheese	non (raw)	Western	red wine, young, rich of tannin		
				red wine, mature, from the Rhone Valley		
				Barolo, mature		
				white wine, dry, fruity		
35	goat cheese	casserole	Western	red wine, velvet, low tannin		
				Burgundy, mature		
				Brunello, mature		
				white wine, dry, fruity		
36	goat cheese	deep fried	Western	red wine, fruity, low tannin, less compound		
				Bordeaux, mature		
				Brunello, mature		
				white wine, dry, fruity		
37	blue mold cheese	non (raw)	Western	Bordeaux, mature		
				Barolo, mature		
				any wine besides smooth one		

				Sauternes		
38	blue mold cheese	casserole	Western	red wine, young, rich of tannin		
				Bordeaux, mature		
				any wine besides smooth one		
				Sauternes		
39	blue mold cheese	deep fried	Western	red wine, dark, fruity, from the “new world”		
				Bordeaux, mature		
				any wine besides smooth one		
				Sauternes		
40	fruit dessert	non (raw)	any	white wine, light, fresh, low acid		
				Gewürztraminer		
				Riesling, rich of acid		
				Port wine		
41	aromatic dessert	non (raw)	any	Gewürztraminer		
				Riesling, semi dry		
				Riesling, rich of acid		
				Port wine		
42	ice cream	non (raw)	any	Port wine		

Additionally we’d like to have your ratings to

5	beef	boiled	Asian	Savignon Blanc		
8	beef	stewed	any	red wine, dark, fruity, from the “new world		

10.7 Questionnaire of the 4th test case solution session
in the field of suggesting an appropriate wine to a given meal

Dear participant,

again, you are asked to suggest an appropriate wine to a given meal.
- If your suggestion is an item of the following list, please, fill in the number in the questionnaire table.
- In case you suggest a different one, please, describe it verbally.

Additionally, you have the opportunity to express that you really can not suggest a wine, because of not having an idea by filling in "*no solution*".

Since some meal descriptions raised some questions, the following specification of some terms might be helpful to interpret it:
- ***Fish*** means "white fish", not red one.
- ***Cheese*** means the one with the strongest taste in its category, because cheese with a light taste would not be a "main ingredient".
- ***Asian style*** means the more spicy variant of it. This is the more Chinese style, less the Japanese one.
- ***Western style*** means no spice, a little salty. Meat in this style goes along with brown sauce (not white one) and in case of fish it is supplemented by garlic.
- ***Any style*** means no spice besides salt and pepper. In case stewed food it also means a dark colored sauce.

Note, that there is a new item in the list – see item #25.

List of possible suggestions

1.	*Red wine, fruity, low tannin, less compound*
2.	*Red wine, young, rich of tannin*
3.	*Red wine, dark, fruity, from the „new world"*
4.	*Red wine, maturely, from the Rhone valley (France)*
5.	*Red wine, velvet, low tannin*
6.	*Pinot Noire*
7.	*Amarone*
8.	*Burgundy, mature*
9.	*Bordeaux, mature*
10.	*Barolo, mature*
11.	*Brunello, mature*
12.	*Beaujolais*

13.	*Rosé, dry, fruity, low tannin*
14.	*White wine, light, fresh, low acid*
15.	*White wine, strong, low tannin*
16.	*White wine, rich in content*
17.	*White wine, dry, fruity*
18.	*Muscatel*
19	*Gewürztraminer*
20.	*Sauvignon Blanc*
21.	*Riesling, semi dry*
22.	*Riesling, rich of acid*
23.	*Port wine*
24.	*Any wine besides smooth one*
25.	*Sauternes*

Again, we encourage you to suggest another kind of wine than listed up above!

To compute more general rules, the meals are described rather general, i.e. not as on a menu card of a restaurant.

Meal #	main ingredient	kind of preparation	style of preparation	suggested wine *# of the above list **or other nomination** (latter is very welcome!)*
1	pork	boiled	Asian	
2	pork	grilled	any	
4	pork	stewed	any	
5	beef	boiled	Asian	
7	beef	fried	any	
8	beef	stewed	any	
10	veal	grilled	any	
11	veal	fried	any	
13	venison	boiled	any	
14	venison	grilled	any	
16	venison	stewed	any	
17	fowl	boiled	any	
19	fowl	fried	any	
20	fowl	stewed	any	
22	fish	steamed	Western	
23	fish	boiled	Asian	
25	fish	fried	any	
26	fish	stewed	Asian	
28	hard cheese	non (raw)	Western	
29	hard cheese	casserole	Western	
31	soft cheese	non (raw)	Western	
32	soft cheese	casserole	Western	
34	goat cheese	non (raw)	Western	
35	goat cheese	casserole	Western	
37	blue mold cheese	non (raw)	Western	
38	blue mold cheese	casserole	Western	
40	fruit dessert	non (raw)	any	
41	aromatic	non (raw)	any	

10.8 Questionnaire of the 4th test case rating session
in the field of suggesting an appropriate wine to a given meal

Dear participant,

thank you very much for your kind suggestions for appropriate wines to given meals. In this **very last** session, we'd like know your rating of other peoples' suggestions to the same cases.

In the column "***correct?***" you can express your agreement to the considered suggestion by filling in "yes" or "no". In the column "***sure?***" you can express some latent internal doubt in this rating by filling in "yes" or "no".

Additionally, you have the opportunity to express, that you can not provide any rating to this solution by filling in "no rating".

Meal #	main ingredient	kind of preparation	style of preparation	suggested wine	correct?	sure?
1	pork	boiled	Asian	red wine, young, rich of tannin		
				red wine, mature, from the Rhone valley		
				Pinot Noire		
				white wine, dry, fruity		
				Muscatel		
2	pork	grilled	any	red wine, young, rich of tannin		
				red wine, velvet, low tannin		
				Amarone		
				white wine, strong, low tannin		
4	pork	stewed	any	red wine, young, rich of tannin		
				Barolo, mature		
				Sauvignon Blanc		
5	beef	boiled	Asian	red wine, dark, fruity, from the "new world"		
				red wine, velvet, low tannin		
				Burgundy, mature		
				Muscatel		

7	beef	fried	any	red wine, young, rich of tannin		
				red wine, dark, fruity, from the “new world”		
				Pinot Noire		
				Burgundy, mature		
				Barolo, mature		
8	beef	stewed	any	red wine, dark, fruity, from the “new world”		
				red wine, mature, from the Rhone valley		
				Bordeaux, mature		
				Barolo, mature		
10	veal	grilled	any	red wine, fruity, low tannin, less compound		
				red wine, young, rich of tannin		
				Bordeaux, mature		
11	veal	fried	any	red wine, fruity, low tannin, less compound		
				red wine, dark, fruity, from the “new world”		
				Burgundy, mature		
				Brunello, mature		
				Rosé, dry, fruity, low tannin		
13	venison	boiled	any	red wine, dark, fruity, from the “new world”		
				Burgundy, mature		
				Bordeaux, mature		
				Barolo, mature		
				Beaujolais		
14	venison	grilled	any	red wine, dark, fruity, from the “new world”		
				Bordeaux, mature		
16	venison	stewed	any	red wine, dark, fruity, from the “new world”		
				Amarone		

				Burgundy, mature		
				Bordeaux, mature		
				Brunello, mature		
17	fowl	boiled	any	red wine, fruity, low tannin, less compound		
				Pinot Noire		
				white wine, light, fresh, low acid		
				Sauvignon Blanc		
19	fowl	fried	any	red wine, fruity, low tannin, less compound		
				red wine, dark, fruity, from the “new world”		
				Beaujolais		
				white wine, dry, fruity		
20	fowl	stewed	any	red wine, fruity, low tannin, less compound		
				red wine, mature, from the Rhone valley		
				Beaujolais		
				white wine, rich in content		
				Riesling, semi dry		
22	fish	steamed	Western	white wine, light, fresh, low acid		
				white wine, dry, fruity		
				Sauvignon Blanc		
23	fish	boiled	Asian	Rosé, dry, fruity, low tannin		
				white wine, light, fresh, low acid		
				white wine, strong, low tannin		
				white wine, rich in content		
				Muscatel		
				Sauvignon Blanc		

25	fish	fried	any	white wine, strong, low tannin		
				white wine, dry, fruity		
				Sauvignon Blanc		
				Riesling, rich of acid		
26	fish	stewed	Asian	white wine, rich in content		
				Muscatel		
				Riesling, semi dry		
28	hard cheese	non (raw)	Western	red wine, young, rich of tannin		
				Pinot Noire		
				Burgundy, mature		
				Bordeaux, mature		
				white wine, rich in content		
29	hard cheese	casserole	Western	red wine, dark, fruity, from the “new world”		
				Burgundy, mature		
				Beaujolais		
				white wine, rich in content		
31	soft cheese	non (raw)	Western	red wine, young, rich of tannin		
				Amarone		
				Burgundy, mature		
				Bordeaux, mature		
				white wine, rich in content		
32	soft cheese	casserole	Western	red wine, dark, fruity, from the “new world”		
				red wine, mature, from the Rhone valley		
				Burgundy, mature		
				Bordeaux, mature		
				white wine, rich in content		
34	goat cheese	non (raw)	Western	red wine, young, rich of tannin		

				red wine, mature, from the Rhone valley		
				Burgundy, mature		
				Bordeaux, mature		
				Barolo, mature		
				white wine, dry, fruity		
35	goat cheese	casserole	Western	red wine, velvet, low tannin		
				Burgundy, mature		
				Bordeaux, mature		
				Brunello, mature		
				white wine, dry, fruity		
37	blue mold cheese	non (raw)	Western	Burgundy, mature		
				Bordeaux, mature		
				Barolo, mature		
				any wine besides smooth one		
				Sauternes		
38	blue mold cheese	casserole	Western	red wine, mature, from the Rhone valley		
				Burgundy, mature		
				Bordeaux, mature		
				any wine besides smooth one		
				Sauternes		
40	fruit dessert	non (raw)	any	Muscatel		
				Gewürztraminer		
				Riesling, rich of acid		
				Port wine		
41	aromatic dessert	non (raw)	any	white wine, light, fresh, low acid		
				Gewürztraminer		
				Port wine		

Thank you very much for your cooperation and patience!